Friedrich-Schiller-Universität Jena
Biologisch-Pharmazeutische Fakultät
Institut für Allgemeine Botanik and Pflanzenphysiologie
Lehrstuhl für Pflanzenphysiologie

Leipniz-Institut für Pflanzengenetik und Kulturpflanzenforschung
Abteilung: Cytogenetik und Genomanalyse
Arbeitgruppe: Karyotypenevolution

In vitro expression, purification and biochemical characterization of KNL2 protein of *Arabidopsis thaliana*

Diplomarbeit
zur Erlangung des Grades eines
Diplom-Biochemikers
vorgelegt von

Michael Sandmann
aus Gera

Jena, 28.06.2013

Gutachter:

Dr. Inna Lermontova

Prof. Dr. Ralf Oelmüller

Abstract

Living cells are propagating by cell division. In eukaryotes, replicated chromosomes are separated by the microtubule apparatus, which attaches to the chromosomes via the kinetochore. The kinetochore is a large protein complex anchored at specific sites of the chromosomes: the centromeres. In many eukaryotes including human and *Arabidopsis*, the centromere is not exclusively determined by its sequence. Instead, the centromere-specific variant of histone H3 (CenH3) is regarded as the central molecule for centromere identity. CenH3 deposition is regulated by a variety of proteins. For instance, the human protein KNL2 is localized at centromeres shortly before CenH3 and mediates CenH3 loading by modulating the chromatin state of the centromeres and recruiting CenH3 loading factors.

In this study the KNL2 homologue of *Arabidopsis thaliana* was characterized. Recombinant *Arabidopsis* full-length KNL2 protein and its N- and C-terminal parts were purified and used as substrates for a kinase assay, for a DNA binding assay and for the production of antibodies against the C-terminal part. The antibodies were used for immunostaining and Western blot experiments. The obtained data showed that *Arabidopsis* KNL2 localizes to the chromocenters during interphase, but not during mitosis. Similar localization was observed by Lermontova et al. (2013) using *KNL2-EYFP* fusion constructs stably transformed in *Arabidopsis*. *KNL2* SALK mutants showed absence of KNL2 immunosignals at chromocenters and specific Western blot signals compared to wild type. *Arabidopsis* KNL2 is phosphorylated by Aurora kinase 3. This phosphorylation is possibly responsible for the absence of KNL2 at chromocenters during mitosis. Furthermore, a DNA binding capability of the C-terminal part is indicated. These observations match with current data in human KNL2 (Maddox, 2013).

Zusammenfassung

Zellen vermehren sich durch Zellteilung. Replizierte Chromosomen werden in Eukaryoten durch eine koordinierte Bewegung von Mikrotubuli getrennt. Dabei sind die Chromosomen über den Kinetochorapparat mit den Mikrotubuli verbunden. Der Kinetochorapparat besteht aus einem Multi-Protein-Komplex, welcher auf bestimmten Chromosomenabschnitten, den Zentromeren, verankert ist. Die Identität der Zentromere wird in vielen Eukaryoten, unter anderen bei Menschen und *Arabidopsis*, nicht notwendigerweise durch eine bestimmte DNS-Sequenz determiniert. Stattdessen wird die zentromerische Variante von Histon H3 (CenH3) als zentrales Molekül angesehen, welches die Identität des Zentromers vermittelt. Die Beladung der Zentromere mit CenH3 ist über eine Vielzahl von Proteinen geregelt. Eines dieser Proteine ist das bei Menschen identifizierte KNL2. Kurz vor der Beladung von CenH3 lokalisiert es am Zentromer, moduliert das zentromerische Chromatin und rekrutiert weitere CenH3 Beladungsfaktoren.
Gegenstand dieser Arbeit ist die Charakterisierung eines homologen Proteins vom humanen KNL2 aus *Arabidopsis thaliana*. Rekombinantes KNL2 Gesamtprotein, sowie N- und C-terminale Fragmente, wurden aufgereinigt und anschließend als Substrate für Kinase-Tests, DNS-Bindingstests und für die Herstellung von Antikörper gegen den C-terminalen Teil benutzt. Die Antikörper dienten zur Durchführung von Immunfärbung und Western Blot Experimenten. Die Ergebnisse dieser Experimente zeigten, dass KNL2 aus *Arabidopsis thaliana* sich in den Chromozentren der Kerne anreichert. Dies geschieht während der Interphase, jedoch nicht in der Mitose. Stabil transformierte *KNL2-EYFP* Konstrukte in *Arabidopsis* zeigten ähnliche Lokalisationen auf (Lermontova et al., 2013). Weder die chromozentrischen Signale noch KNL2-spezifische Western Blot Signale wurden in Proben von *KNL2* SALK Mutanten gefunden. Ein Kinase-Test zeigte, dass *Arabidopsis*-KNL2 von Aurora Kinase 3 phosphoryliert wird. Diese Phosphorylierung könnte ein Grund für die Abwesenheit von KNL2 während der Mitose darstellen. Des Weiteren liefern die Ergebnisse dieser Arbeit Indizien für eine DNS-Bindefähigkeit im C-terminalen Teil von KNL2. Letzteres findet sich auch im humanen KNL2 Protein (Maddox, 2013).

Table of contents

List of figures

List of tables

List of abbreviations

Common abbreviations of nucleic acids and scientific units are not shown.

Abbreviation	**Full word/Explanation**
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AIM	*Agrobacterium* induction medium
APS	Ammonium persulfate
attB (sites)	Gateway® term
BiFC	Bimolecular fluorescence complementation
bp	Base pairs
BP (clonase/reaction)	Gateway® term
BSA	Bovine serum albumine
C. elegans	*Caenorhabditis elegans*
CAF	Chromatin assembly factor
CATD	CENPA targeting domain
cDNA	Complementary DNA
CDP/Cux	CCAAT displacement protein/Cut homeobox transcription factor
cen180/pAL	180 bp repeat in *Arabidopsis thaliana* centromeres
CenH3	Centromeric histone 3
CENPA	Centromere protein A
ChIP	Chromatin immunoprecipitation
CID	Centromere identifier protein
Clp	Caseinolytic protease
Col-0	Columbia 0
CpG	Cytosin phosphat guanine
Cse4/Cse4p	Chromosome segregation protein 4
DAPI	4',6-Diamidino-2-phenylindole
DBD	DNA binding domain
dcp	Dipeptidyl carboxypeptidase
DDM1	Decrease in DNA methylation 1
DIC	Differential interference contrast
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
Ed	Editor

EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
EYFP	Enhanced yellow fluorescent protein
G_1 phase	Gap 1 phase
G_2 phase	Gap 2 phase
gDNA	Genomic DNA
GFP	Green fluorescent protein
H3K4me2	Dimethylation on lysine at position 4 of histone H3
H3K9me2	Dimethylation on lysine at position 9 of histone H3
H3K9me3	Trimethylation on lysine at position 9 of histone H3
H3S10ph	Phosphorylation of serine 10 of histone H3
H3S28ph	Phosphorylation of serine 28 of histone H3
HCCA	α-Cyano-4-hydroxycinnamic acid
HFD	Histone fold domain
His tag	6x Histidin tagged
HJURP	Holiday junction recognition protein
hMis	Human Mis
HRP	Horseradish peroxidase
HsKNL2	*Homo sapiens* KNL2
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Isl1	Islet 1
KNL2	Kinetochore-null 2
KNL2 variants	KNL2, KNL2Nterm and KNL2Cterm
KNL2Cterm/KNL2C	C-terminal part of KNL2
KNL2Nterm/KNL2N	N-terminal part of KNL2
LB	Lysogeny broth
LR (clonase/reaction)	Gateway® term
MALDI-TOF	Matrix assisted laser desorption/ionization-time of flight
MBP	Maltose binding protein
MES	2-(*N*-Morpholino) ethanesulfonic acid
Mis18BP1	Mis18 binding protein 1
MS	Mass spectrometry

MTSB	Microtubules stabilizing buffer
MWCO	Molecular weight cut off
NTA	Nitrilotriacetic acid
OD	Optical density/absorbance
PAA	Polyacrylamide
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pI	Isoelectric point
PIPES	Piperazine-N, N′-bis (2-ethanesulfonic acid)
PMSF	Phenylmethanesulfonylfluoride
poly dI/dC	poly-deoxyinosinic-deoxycytidylic acid sodium salt
PPB	Preprophase band
PVDF	Polyvinylidene fluoride
RNAi	RNA interference
rpm	Rounds per minute
RT	Room temperature
RT-PCR	Reverse transcriptase PCR
S phase	Synthesis phase
SALK	Salk institute genomic analysis laboratory
SANT	Swi3, Ada2, N-Cor, and TFIIIB
SANTA	SANT associated
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolite repression
SR-SIM	Super resolution-structured illumination microscopy
TBE	Tris/borate/EDTA
TBS	Tris-buffered saline
TE	Tris EDTA
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
Tris	Tris (hydroxymethyl) aminomethane
YEB	Yeast extract and beef
YFP	Yellow fluorescent protein

1 Introduction

1.1 Cell cycle in eukaryotic organisms

Living cells reproduce themselves to maintain their existence over million of generations. This could be done by an asexual or sexual process (Lodish et al., 2000).

1.1.1 Asexual cell division

A case of asexual reproduction: One eukaryotic 'parent' cell generates two daughter cells with the genome of the parent in a cell cycle. The cell cycle consists of 4 phases: the S phase, in which the DNA is replicated, the M phase, in which the two replicates – the sister chromatids – and the cytoplasm are separated, and two gap phases (G_1 and G_2), which connects M and S phase. G_1, S and G_2 together are also known as interphase (Cooper, 2000).

1.1.1.1 Interphase of the cell cycle

G_1 is the gap phase from M to S phase. S phase is a period of DNA replication. The replicated sister chromatids join together at their centromeres up to the beginning of the anaphase (Alberts et al., 2002). The centromeres are defined as the primary constriction of the chromosomes (see 1.2) (De Rop et al., 2012). In mammalians, the S phase occupies about the half of the cell cycle (Alberts et al., 2002). G_2 is the connection between S to M phase. G_1 and G_2 phase are growth phases, in which the cell synthesize protein and increase the number of organelles to compensate cytoplasmic division and prepare for the replication and separation of the genome. Additionally cells monitor the environment during these phases. Especially unfavorable conditions during G_1 lead the cell to enter the specialized resting phase G_0, which can be switch back to G_1 later under suitable conditions (Alberts et al., 2002).

1.1.1.2 M phase of the cell cycle

M phase is divided into sub-phases: mitosis or nuclear division, and cytokinesis or cytoplasmic division (Alberts et al., 2002).

The mitosis can be subdivided in 5 stages: prophase, prometaphase, metaphase, anaphase and telophase. In prophase, the chromosomes begin to condense and the formation of the mitotic spindle apparatus with two poles starts. In the next stage, nuclear envelope breaks down, the attachment of the spindle apparatus to chromosomes via kinetochores begins and

the chromosome movement starts. The kinetochore is a molecular apparatus, which consists of DNA and proteins and maintains chromosome segregation (Karpen and Allshire, 1997; Sullivan et al., 2001). In metaphase, the kinetochore of each sister chromatid attach to spindle apparatus and the chromosomes are aligned between the spindle poles in an equatorial orientation. During anaphase spindle apparatus mediates segregation of sister chromatids. In the last stage, the chromatids arrive at the poles and decondense, two new nuclear envelopes form around the two separated sets of sister chromatids, as well as the contractile ring assembles, which herald the cytokinesis (Alberts et al., 2002).

The contractile ring divides the cytoplasm during cytokinesis. This lead to the generation of two daughter cells and it completes the asexual cell division (Alberts et al., 2002).

1.1.2 Sexual cell division

The eukaryotic sexual reproduction cycle consists of the fusion of two haploid cells to a diploid one and the division of a diploid cell in two haploids by meiosis. Haploid cells require a reduced number of homologue chromosomes; otherwise, the chromosome number will multiple with every fusion. Meiosis achieves the haploid level of a replicated, diploid genome by a two step process: the division I of meiosis, which is a separation of the homologue chromosomes; and the division II of meiosis, which resembles mitosis. Therefore, the separation of the sister chromatids occurs in division II (Lodish et al., 2000; Alberts et al., 2002). Haploid or diploid cells can undergo mitosis and maintain the proliferation depending on the species. In animals, the diploid state is the preferred state for proliferation, in fission yeast it is the haploid state and plants show either one of both states (Alberts et al., 2002).

1.2 Centromeres and centromere identity

A coordinated separation of chromosomes is crucial in mitosis and meiosis. The centromeres play an important role in this process. Centromeres are the primary constriction of metaphase chromosomes and consist of DNA and protein parts (van der Ploeg et al., 1974; Sullivan et al., 1996; Fang and Spector, 2005). The centromeres of plants and animals contain regions of highly repetitive satellite DNA (Henikoff et al., 2001). Typical centromeric repeats of *Arabidopsis thaliana* are 500-1000 kb regions with a characteristic length of 180 bp (referred as cen180). Moreover, transcribed regions like 5s rDNA or genes like the phosphoenolpyruvate phosphate translocator are found in

Arabidopsis centromeres. The centromeric sequences have a decreased level of recombination in *Arabidopsis* (Copenhaver et al., 1999).

The centromeres are the chromosomal positions, where the kinetochore assembles (Lermontova and Schubert, 2013). The kinetochore consists of the inner kinetochore, a protein network attached to DNA, which are necessary for kinotochore assembly; and the outer kinetochore, which maintains microtubule binding (Pluta et al., 1995). The question arise how the centromere is recognized by the kinetochore and how the centromere is identified?

The sequence of centromeres is not regarded to maintain solely centromere function or identity in higher eukaryotic organism (Warburton et al., 1997; Henikoff et al., 2001; Nasuda et al., 2005). In contrast, the 'point centromere' and its sequence of *Saccharomyces cerevisiae* is known to be necessary for proper mitosis (Pluta et al., 1995; Sullivan et al., 2001). "The primary sequence of centromeric DNA is not conserved among species, or even among different chromosomes of individual organism."(Karpen and Allshire, 1997) The appearance of higher ratios of adenine/thymine pairs and the embedding in heterochromatic, pericentromeric regions are the known similarities of eukaryotic centromeres (Karpen and Allshire, 1997; Sullivan et al., 2001; Fang and Spector, 2005). The lack of a common, high conserved DNA element in centromeres indicates the existence of other mechanisms mediating kinetochore assembly. The phenomenon of 'neocentromeres' formation, which do not require typical centromeric sequences, supports this indication (Peacock et al., 1981; Brown and Tyler-Smith, 1995; Depinet et al., 1997; Karpen and Allshire, 1997; Warburton et al., 1997). Neocentromeres can carry out normal centromere function and are established on the basis of epigenetic mechanism (Maggert and Karpen, 2000). *In vivo* studies shows that epigenetic changes can activate nonfunctional eukaryotic centromeres (Steiner and Clarke, 1994). Some examples for epigenetic, centromere-characteristic changes, which appear also in *Arabidopsis*, are described below.

1) Centromeres and neocentromeres are hypermethylated at cytosines of CpG dinucleotides. However, some pockets of centromeric sequences are in non-methylated state and show transcriptional activity (Wong et al., 2006). 2) The di- or trimethylation on lysine 9 of histone H3 (H3K9me3/H3K9me2) are the major histone modifications at centromeric sites in maize and *Arabidopsis* (Shi and Dawe, 2006; Kim et al., 2012). 3) Some centromeric transposons and related tandem repeats were transcribed into dsRNA, which were subsequently configured into RNAi molecules by the RNAi machinery. These

RNAi molecules may affect histone 3 lysine 9 methylation, DNA methylation and they are involved in establishment and maintenance of (pericentromeric) heterochromatin (Volpe et al., 2002; Lippman et al., 2004). 4) Some studies refer to a 'higher order structure' of centromeric DNA sequences, which is specifically folded and can attract the kinetochore (Karpen and Allshire, 1997; Sullivan et al., 2001). 5) CenH3, which is a centromeric variant of histone H3, is regarded as central epigenetic mark for centromere identity (Sullivan et al., 2001; Panchenko et al., 2011) (see 0).

In general epigenetic marks can be quite species specific (Shi and Dawe, 2006; Roudier et al., 2011) and although histone modifications are highly conserved, they should not be generalized among species (Fuchs et al., 2006).

1.3 Epigenetic regulation at centromeric and pericentromeric sites in *Arabidopsis*

CenH3 associated core centromeric satellite DNA of *Arabidopsis thaliana* and *Zea mays* are hypomethylated in comparison to their related pericentromic, heterochromatin-enriched regions. The centromeric hypomethylation is correlated with a significant reduced level of H3K9me2 on sequences with a distinct distribution of CG and CNG sites (Zhang et al., 2008). Another study confirms the increased level of H3K9me2 methylation in *Arabidopsis* pericentromeric regions, where the dimethylation forms patterns of long and rarely interrupted blocks (Bernatavichute et al., 2008).

In *Arabidopsis* H3K9me2 typically marks heterochromatin and regions of silenced transposons, whereas the H3K9me3 modification is in general associated with euchromatic sequences (Roudier et al., 2011; Veiseth et al., 2011). However, a locus-specific repression of transposons via an H3K9me3 signal is possible, too (Veiseth et al., 2011). DNA methylation and methylation of H3K9 position is performed by siRNA guided DDM1 (Lippman et al., 2004). Additionally less H3K9me2 modification occurs in maize pericentromeric areas, however, H3K9me1 is enriched (Shi and Dawe, 2006). There are other important sites for epigenetic modifications like H3K4 (Roudier et al., 2011). Nevertheless, the H3K9 position emphasizes the diverse role of epigenetic modifications for centromere identity as one possible example.

CenH3 and its role in eukaryotic organisms

CenH3 (alternative name: Htr12) in *Arabidopsis*, his mammalian equivalent CENPA, CID in fly or Cse4 in yeast replaces histone H3 in centromeric nucleosomes (Shelby et al., 1997; Meluh et al., 1998; Sullivan et al., 2001). Classical octameric nucleosomes consist of dimers of histone H2A, histone H2B, histone H3 and histone H4 (Nechemia-Arbely et al., 2012). Centromeric histone variant CenH3 could be seen as "evolutionary link between the seemingly divergent centromeres" (Sullivan et al., 2001) in eukaryotic organism. CenH3 can carry out the transformation of non-centromeric regions to functional neocentromeres (Lo et al., 2001). In artificial LacI-*LacO* experiments, CenH3 is able to establish functional centromeres at predefined, stably integrated sites inside the *Drosophila* and *Arabidopsis* genome (Mendiburo et al., 2011; Teo et al., 2013). These findings show a self-propagating and heritable character of CenH3 as epigenetic mark, too. Furthermore, CenH3 can mediate higher-order chromatin structure, "modifies the entry/exit of the DNA from the nucleosome" (Nechemia-Arbely et al., 2012) and possibly change the packaging of nucleosome bound DNA (Tachiwana et al., 2011).

The incorporation of CenH3 can be regarded as hallmark for centromere identity. Blocks of CenH3 together with blocks of histone H3 form the centromeric chromatin (Zhang et al., 2008).

Inactivation of *CID* leads to disrupted mitosis and cell-cycle progression in fly (Blower and Karpen, 2001) and *Arabidopsis CenH3* knock-down causes dwarfism connected to a reduced number of mitosis as well as a reduced fertility due to disordered meiosis (Lermontova et al., 2011). Similar results can be found in other species (i.e.: mice (Howman et al., 2000), yeast (Meluh et al., 1998) or nematode (Oegema et al., 2001)). Moreover, these studies can also show that the proper localization of other kinetochore proteins depends on CenH3.

1.3.1 Function of the N- and C-terminus of CenH3

CenH3 homologues are less conserved among species than histone H3 despite of conserved function. This is especially true for the N-terminal tail and loop 1 of the C-terminal region, which is not conserved (Henikoff et al., 2001; Cooper and Henikoff, 2004). The C-terminal region of CenH3 contains the HFD domain, which is responsible for centromere specific DNA binding in eukaryotic organisms. The non-conserved loop 1 region and the adjacent α2-helix of the HFD forms the CATD domain (Shelby et al., 1997; Lermontova et al., 2006; Perpelescu and Fukagawa, 2011; Nechemia-Arbely et al., 2012).

The “loop 1 in the CATD protrudes from the CENP-A nucleosome” (Nechemia-Arbely et al., 2012) and may attract some trans-acting centromeric compounds (Tachiwana et al., 2011). The substitution of three certain, among plants conserved amino acids of the loop 1 domain leads to abolished centromere targeting (Moraes et al., 2011). Therefore, these amino acids are necessary for centromere recognition and the loop 1 domain contains certain conserved and functional spots or maybe an adaptive evolution process occurs between centromeres and CenH3 is the basis on variable N-terminus or loop 1 of the CATD (Cooper and Henikoff, 2004).

The role of the N-terminus seems to be divergent among species. In *Saccharomyces cerevisiae*, the N-terminus is involved in protein-protein interaction at the kinetochore (Chen et al., 2000). However, histone H3 containing CATD chimers can localize at centromeres and rescue growth of CENPA deficient human HeLa cells. Thus, the N-terminus seems not to be necessary for kinetochore assembly in human (Black et al., 2007). In *Arabidopsis*, the role of N-terminus of CenH3 differs between mitosis and meiosis. In mitosis, the C-terminal part is sufficient, however, for proper meiosis the N-terminal part of CenH3 is also required (Lermontova et al., 2011; Ravi et al., 2011; Lermontova and Schubert, 2013).

1.4 CenH3 assembly in *Arabidopsis*

CenH3 localizes at cen180 positions of centromeres, but not at elements like *Athila* retrotransposons, which are especially enriched in pericentromeric regions (Nagaki et al., 2003; Fang and Spector, 2005; Lermontova and Schubert, 2013). However, CenH3 proteins are not loaded on the entire cen180 repeats. Chromatin fiber immunolabeling showed a loading of 12%-15% of cen180 repeats by CenH3 at the knobs on the extended chromatin fibers (Shibata and Murata, 2004; Lermontova and Schubert, 2013). The loading of CenH3 in *Arabidopsis* centromeres is happening mainly during G_2 phase (Lermontova et al., 2006). In other phylogenetic groups, the loading of centromeric histones occur in different cell cycle phases (Lermontova et al., 2006; Nechemia-Arbely et al., 2012; Lermontova and Schubert, 2013). The reason behind this difference is unclear so far, maybe the non-conserved parts of CenH3 play a role.

The incorporation of CenH3 at centromere is shown as essential for centromere identity and function. Therefore, the questions arise how CenH3 is loaded and which components are involved in loading mechanism?

1.4.1 Models of CenH3 loading

There are different models of CenH3 loading/propagation: last to replicate model, or in contrast, the early to replicate model, nuclear organization model and cyclical chromatin model (Sullivan et al., 2001).

The last/early replicate model can not be applied, because CenH3 assembly has to happen in S phase according to this model. The nuclear organization model requires distinct centromeric domains, where CenH3 assembly occurs at clustered centromeres. The *Arabidopsis* centromeres are usually distributed at the nuclear membrane, but do not cluster together (Lermontova et al., 2006). The cyclical chromatin assembly model relies on CAFs. This model might explain the assembly of new CenH3 histones in *Arabidopsis*. CenH3 and H3 nucleosomes are dispersively segregated on daughter strands during replication (Annunziato, 2005). CAFs replenish the free nucleosome sites according to remaining neighbor-nucleosomes. The assembly of histone H3 is dependant on S phase, whereas CenH3 not (Shelby et al., 2000; Sullivan et al., 2001). In human, HJURP is such a CAF, a histone chaperone, which can induce the formation of neocentromeres (see 1.2) (Barnhart et al., 2011).

There are some models, which explain how the different loading of histone H3 and CenH3 is organized. CAFs can possibly exchange CenH3 against 'placeholder histones' after histone H3 loading (Sullivan, 2001; Dunleavy et al., 2011). Also, the pericentromeric heterochromatin may block conventional nucleosome assembly. Therefore, later CenH3 assembly is possible (Ahmad and Henikoff, 2001). Other findings in fly, human and yeast show that CenH3 can be present in tetrameric instead of classical octameric nucleosomes (Dalal et al., 2007; Bui et al., 2012; De Rop et al., 2012). During replication the octameric CenH3 nucleosomes might be split in two and divided on the daughter strands. This arises as a possible mechanism for an epigenetic inheritance throughout the whole cell cycle (Dalal et al., 2007; Bui et al., 2012). CAFs could hypothetically replenish octameric CenH3 nucleosomes in G_2 phase (Lermontova et al., 2011).

1.4.2 CenH3 licensing and maintenance

The process of CenH3 loading can be subdivided into three parts: licensing, loading or deposition and maintenance (Lermontova et al., 2013). 1) In human and nematodes the Mis18 complex with its component KNL2 is annotated for the function of licensing and recognition (Fujita et al., 2007; Maddox et al., 2007; De Rop et al., 2012). This protein and its function are discussed in the next chapter. 2) The loading of CenH3 is shown in the previous chapter. 3) The maintenance of the newly incorporated CenH3 histones can be

carried by the GTPase activating human protein MgcRacGAP (Lagana et al., 2010; De Rop et al., 2012; Nechemia-Arbely et al., 2012). This protein can bind hsKNL2 and might activate other proteins, which are necessary for CenH3 maintenance and are recruited by hsKNL2 (Lagana et al., 2010).

1.5 KNL2 homologues in other organisms

C. elegans or human KNL2 have a C-terminal Myb/SANT domain (Maddox et al., 2007). The Myb domain is annotated with DNA binding (Lipsick, 1996) and the SANT domain binds modified histone tails or other modulating proteins and plays often a general role in chromatin dynamics and remodeling (Boyer et al., 2004). Myb and SANT domain shares high similarity on primary sequence of the DBD and secondary structure. However, SANT domain do not contain several conserved DNA-binding specific key residues of Myb DBD and is functional divergent to this similar domain (Boyer et al., 2004). Recent studies show a motif-independent DNA binding capability of nematode/human KNL2 (Maddox, 2013). Therefore, these homologues are annotated by a C-terminal Myb domain.

In silico analysis can identify a SANTA domain in the N-terminal part of *Arabidopsis KNL2* (Lermontova et al., 2013). SANTA domain could carry out protein-protein interaction, is associated with SANT domains and possibly regulates chromatin remodeling as well. Although SANTA and SANT are separate domains, there are functionally related and evolutionary linked (Zhang et al., 2006). Moreover, to the similarity of Myb and SANT domain, hints are gives, that our candidate and human or nematode KNL2 are related homologues. However, question about the functional background are still open.

Homologues of KNL2 show an important role in CenH3 loading. "[C]entromeric localization [of mice KNL2] is essential for the subsequent recruitment of de-novo synthesized CENPA." (Fujita et al., 2007) RNAi knock down of mice *KNL2* shows abortion of CENPA loading and a disturbed mitosis (Fujita et al., 2007). Homologues of holocentric *C. elegans* or monocentric human show similar function and similar results for RNAi knock-down (Maddox et al., 2007). *KNL2* depletion mimics *CenH3* depletion. The loading of HsKNL2 happens during anaphase/mitotic exit to early G_1 and CENPA loading occurs in late telophase/G_1 phase in human HeLa cells (Maddox et al., 2007; Nechemia-Arbely et al., 2012). So KNL2 seem to preludes CenH3.

The HsKNL2 complex (Mis18) possibly plays a role in histone 4 acetylation, recruiting of DNA methyltransferases, H3K9me3/H3K9me2 or H3K4me2 modification, and it is

required for the recruitment of HJURP to endogenous centromeres (see 1.4.1). Therefore, HsKNL2 change the epigenetic mark of centromeres and 'priming' CenH3 loading (Barnhart et al., 2011; Perpelescu and Fukagawa, 2011).
Additionally the primary sequence of KNL2 homologues contains certain conserved motifs, however, they differ in primary sequence (Maddox et al., 2007; Lermontova and Schubert, 2013).

1.6 The regulation of KNL2 function

1.6.1 Interaction between CenH3 and KNL2

A direct interaction between CenH3 and HsKNL2 complex could not be found (Fujita et al., 2007; De Rop et al., 2012). However, a recent study shows that CENPC binds directly HsKNL2 and it is involved in HsKNL2 recruiting at centromeres (Moree et al., 2011). CENPC is part of the inner kinetochore (see 1.2) and binds CENPA. Therefore, CENPC provides a foundation for a CENPA-HsKNL2 network (Carroll et al., 2010). The function of KNL2 might not be relying on a close interaction between CenH3 and KNL2.

1.6.2 Regulation of KNL2 by phosphorylation

Additionally function of HsKNL2 depends on phosphorylation. HsKNL2 targets centromeres only in unphosphorylated state, furthermore, de-phosphorylation and deactivation of the related kinase happens at the loading time of HsKNL2 (Nechemia-Arbely et al., 2012).

1.7 Aim of the study

In this study we want to produce recombinant *Arabidopsis* KNL2 in full length and as well as N- and C-terminal parts of it. The recombinant proteins are used for three purposes.
Recombinant KNL2 shall be used as substrate in a kinase assay with Aurora kinases and as antigen for antibody production in rabbit. The antibodies should be used in immunostaining assays and CenH3-KNL2 co-localization experiments. Additionally we performed EMSA to look for a DNA binding capability like HsKNL2 (see 1.5). These preliminary experiments should give some insights for an uncharacterized *Arabidopsis* protein.

2 Methods

2.1 Gateway based gene cloning

For the expression of KNL2 protein and its N- and C-terminal parts (=KNL2 variants) in *E. coli* entry vectors were generated using the Gateway cloning technology (Life technologies®). cDNAs of *KNL2* variants were amplified by PCR and cloned into the pDONR221 vector using BP clonase reaction (see 2.5, Supplemental Figure 1 & Lermontova et al. (2013)). The purification of attB – PCR products and the BP reaction were done according to Life technologies®'s manual. The BP reaction was set up according to the following scheme: 1-2 µl purified PCR product (20-50 fmol), 1 µl pDONR™221 (150 ng/µl, Life technologies®, 12536-017) and 2 µl BP clonase (Life technologies®); the total volume of the reaction was adjusted to 8 µl by TE buffer. After 1h of incubation at room temperature the reaction was terminated by addition of 1 µl of proteinase K. To desalt samples and prepare them for transformation of *E. coli* by electroporation (see 2.2) a dialysis step was performed for 30 min on a filter membrane (Millipore, nitrocellulose filter 0.025 µm, #VSWP02500) against distilled water. pDONR221 vectors carrying the *KNL2* variants were checked by PCR and sequencing (data not shown subsequently). Generated entry vectors were used for the recombination into the expression vector pDEST17 using the LR clonase reaction (Life technologies®) according to manufactures instructions with slight modifications. The incubation of the LR clonase reaction was reduced to 22°C. A nitrocellulose filter dialysis step was also carried out. The DNA concentration of the dialyzed LR clonase reaction was measured by nanodrop (see 2.7). The pDEST17, which contained the sequence of *KNL2* variants, were checked by sequencing (see 2.9) and subsequently transformed into *E. coli* BL21 strain by electroporation to express recombinant KNL2 variants (see 2.2 & 2.11).

2.2 Transformation of *Escherichia coli* by electroporation

Dialyzed BP and LR clonase reactions were used for the transformation of *E. coli* strain DH5α to select positive clones. Cloned vectors were isolated from the positive selected cells (see 2.3) to obtain material for subsequent applications (see 2.1).

Deep-frozen electroporation-competent *E. coli* cells (DH5α or BL21) were thawed on ice for 5-10 min with gentle mixing in between. 1-50 ng of dialyzed BP or LR clonase reaction

were added to 50 µl competent cells followed by gentle mixing. Electroporation was done at 2500 V with Easy Jet prima (Equi Bio®). Afterwards 500 µl of SOC medium was added, samples were incubated at 37°C for 1 h. 100 µl-250 µl culture was distributed onto sterile, pre-warmed selective LB-agar plates, which were subsequently incubated for 12 to 16 h at 37°C. Only transformed, positive cells should grow on the selective, antibiotic containing medium.

2.3 Plasmid isolation

Liquid cultures were prepared with bacterial glycerol stocks (see 0), which were diluted 1:100 in 3 to 5 ml sterile LB medium and incubated overnight at 37°C with shaking. QIAGEN®´s QIAprep™ spin Miniprep kit (#27106) was used for the plasmid isolation of liquid cultures according to manufacture's manual.

2.4 Polymerase chain reaction (PCR) for amplification of distinct DNA fragments

PCR mastermix (see 3.1) was prepared. Template DNA was added 1:25 to the mastermix. The isolated (see 2.3) expression vector pDEST17 with *KNL2* variant cDNA inserts was used as template in the experiment of chapter 4.1.1 and diluted (1:10000), isolated *Arabidopsis* genomic DNA (see 2.8) served as template for the generation of cen180 fragments (see 4.2.1). Experiment dependent PCR programs were executed with the PCR machine of BioRad® (iCycler Thermo Cycler). The sequences of the forward and reverse primers and the PCR program are shown in the chapter 3.2.

2.5 PCR sample purification

QIAGEN®´s QIAquick® PCR purification kit protocol was used. Distilled water or 10 mM tris HCl pH 7.5-8.5 served as eluent.

2.6 Agarose gel electrophoresis for DNA analysis

DNA gels were prepared according to Green and Sambrook (2012). In general, 1% agarose were used. Ultrapure agarose (Life technologies, #15510-027) was resuspended in 1x TBE buffer and boiled in microwave. DNA staining substance SYBR® safe (life technologies®,

S33102) was added after the liquid gel cooled down to approximately 60°C. For the loading, of DNA samples 4 volumes were mixed with 1 volume of 5x DNA loading buffer loaded into the polymerized gel. DNA markers (from Thermo Fisher Scientific®: SM0331 and SM0613) were used for the estimation of DNA size of separated fragments. The stained DNA was visualized under a trans-illuminator (Syngene, Gene flash).

2.7 Estimation of DNA concentration

DNA concentration was estimated by measurement of DNA absorption at OD260 on nanodrop machine (prqlab® Biotechnologie´s ND-1000) according to producer's manual.

2.8 Extraction of genomic DNA from plant material

Genomic DNA was isolated from ≤100 mg of leaf material of *Arabidopsis* according to the manufacturer's manual of the plant DNA extraction kit (DNeasy® plant mini kit by QIAGEN®). Plant tissues were disrupted using Retsch®´s MM301 machine. Distilled water was used for the elution of the genomic DNA. The obtained gDNA served as template for the cen180 PCR (see 2.4).

2.9 DNA Sequencing

About 200 ng of isolated plasmid DNA (see 2.3 & 2.7) were used for sequencing by service team of Dr. Patrick Schweizer (IPK, Gatersleben). T7 promoter and T7 terminator primers were used (see 3.2).

2.10 Sequence alignment

DNA sequences were analyzed by a multiple sequence alignment tool (EMBL-EBI, Clustal Omega) to compare sequencing data (see 2.9) with annotated cDNA sequences or to compare different database identifiers (Supplemental Figure 8).

Generation of bacterial glycerol stocks

Glycerol stocks were prepared with 3 volumes of liquid bacterial overnight culture (see 2.11) and one volume of glycerol (100%) to store these bacterial cultures over long time. Alternatively the protocol of Green and Sambrook (2012) was used instead.

2.11 Protein expression

E. coli overnight cultures were grown in 1-10 ml of LB medium with addition of ampicillin (final concentration 200 µg/ml), which were inoculated by glycerol stocks (see 2.3 & 0) or single colonies grown on petri dishes with agar LB medium were picked. Liquid cultures were incubated overnight at 37°C with shaking at 180 rpm. 12-16 h old overnight cultures were inoculated in ampicillin containing (200 µg/ml), pre-warmed (37°C) LB medium with a minimum 1:100 dilution. Optionally, the overnight culture could be centrifuged (5000 g, 10 min, 4°C) and resuspended in ampicillin containing LB medium before inoculation. Then, the culture was growing at 37°C with shaking 180 rpm until bacterial culture reached OD600 values between 0.5-1.0 (Eppendorf, BioPhotometer, #6131). Usually OD 0.8 was targeted. Protein expression was induced by IPTG (final concentration) of 1 mM and new ampicillin was added. Samples were taken 3 h after induction and were subsequently centrifuged (5000 rpm, 15 min, 4°C). Supernatants were removed and samples were stored at -80°C until they were analysed by SDS-PAGE (see 2.17) or solubility test (see 2.13), alternatively they were used for protein purification (see 2.16.2).

2.12 Total protein extraction from *E. coli* cells

SDS gel-loading buffer was added to the pellet of induced bacterial culture (see 2.11). The volume of the buffer was 1/25 of the volume of bacterial culture before centrifugation. The resuspended pellet was warmed at 65°C for 15 min. A sonication step (compare 2.13.1) for 1 min was performed to break the cells followed by a centrifugation (13000 rpm, 10 min, RT). The supernatant was heated at 95°C for 5 min. The samples were cooled down and were centrifuged (13000 rpm, 1 min, RT). The validation of the samples was performed by SDS-PAGE (see 2.17).

2.13 Solubility test

2.13.1 Protein solubility analysis

1 ml of bacterial culture (see 2.11) was taken 3h after IPTG induction and centrifuged at 5000 rpm for 15 min. The pellet was dissolved in 100 µl of TE buffer. A suspension was sonicated (Sonics & Materials®, Vibra Cell; 60% power) for 90 s with a schedule of 5 s pulse time and 5 s interruption to destroy the cell wall. The sonicated sample was centrifuged at 13,000 rpm for 15 min. Then, 2 volumes of SDS gel-loading buffer was added to supernatant (soluble fraction) and 100µl of SDS gel-loading buffer dissolved the pellet (insoluble fraction). Subsequently the samples were analyzed by SDS-PAGE (see 2.17).

2.13.2 Solubility test according to Green and Sambrook (2012)

The procedure was performed according to the named source with slight modifications. Benzonase (Benzonase® Nuclease #70746-4 by Novagen®/EMD Millipore) was used instead of DNase I (Thermo Fisher Scientific® #EN0521). The *E. coli* cells were disrupted by sonication in bacterial lysis buffer (compare 2.13.1).

2.14 Protein extraction from *Arabidopsis thaliana* material

100 mg of plant material (young seedlings) was frozen in liquid nitrogen and stored at -80°C or directly homogenized by mechanical disruption (Retsch MM301) for 30s at 30 Hz. The homogenate was resuspended in 200 µl-1 ml of SDS gel-loading buffer. The samples were incubated for 10 min at 70°C with shaking at 800 rpm. A centrifugation step (13000 rpm, 5 min, RT) was carried out and supernatants (protein extracts) were transferred into new micro-reaction tubes.

2.15 Determination of protein concentration by Bradford assay

Proteins were first precipitated by 10% TCA to get rid of SDS and DTT**,** which are not compatible with Bradford solution, to determine protein concentration in protein extracts from *Arabidopsis thaliana* material (see 2.14) in gel-loading buffer.
1 volume of *Arabidopsis* protein extract and 5 volumes of 10% TCA were mixed. Samples were incubated for 30 min on ice. A centrifugation (13000 rpm, 10 min, and 4°C) was

performed. Pellets were resuspended in 100 µl of 0.1 N NaOH. Bradford assay was done according to BioRad®´s Quick Start™ Bradford Protein Assay instruction. 10 or 20 µl of each protein samples dissolved in 0.1 N NaOH were transferred into new micro-reaction tubes. Then, 1 ml of protein staining dye 1x (BioRad® #500-0205) was added. After 5 min incubation at room temperature samples were measured at OD595 by the spectro-meter BioRad® SmartSpec Plus. BSA was used for calibration curve.

Samples gained from protein purification (see 2.16.2) were dialyzed (compare 2.22 or 2.28). The protein samples were compatible to Bradford assay after dialysis. The protein concentrations were used for the calculations of protein-DNA ratios in the EMSA assay (see 2.22).

2.16 Protein purification after protein expression

The recombinant KNL2 proteins, their transcription were induced by IPTG (see 2.11), have to be purified to omit bacterial protein and gain pure recombinant protein for other application like kinase assay, antibody production or EMSA (see 1.7, 2.22, 2.25 & 2.27). The His tag of the recombinant protein, which is encoded on the pDEST17 vector (see 2.1), was used for a specific His tag purification. Two different approaches were applied.

2.16.1 Protein purification with QuickPick™ IMAC PLUS

The manual of the manufacturer 'Bio-Nobile Oy' were used for protein purification under denaturing conditions with one modification. The incubation time of the recombinant proteins in elution buffer was increased to 90 min to increase the yield of eluted, purified protein.

2.16.2 Protein purification with Ni^{2+}-NTA beads under denaturating conditions (batch method)

Agarose beads slurry (Macherey-Nagel®, 745400.25 or QIAGEN®, 1000632) were equilibrated two times in 1x PBS followed by two times equilibration in GuHCl pH 8 buffer. The frozen cell pellets of IPTG induced cultures (see 2.11) were resuspended in GuHCl pH 8 buffer with 1 mM PMSF and sonicated (Sonics & Materials®, Vibra Cell; 60% power) for 1-5 min until the solution get clear. PMSF is a serine proteinase inhibitor, which shall omit protein degradation. The sonicated cell suspensions were incubated for 1 h with shaking. The pH of the suspensions were adjusted to pH 8 followed by a centrifugation (8000 g, 10 min, RT). Supernatants were combined with equilibrated beads.

The mixtures were incubated for 1 h with shaking to let the recombinant His tagged protein bind to the beads. Then, the samples were centrifuged (1000 g, 2 min, RT). Washing steps were performed by using urea containing buffers with pH 8 and pH 6 or optional with pH 5.3 (see 3.1). The protein was eluted by urea containing buffers with pH 4 and pH 2.8. The volume of the elution buffer was equal to the volume of beads slurry. In the case the samples were eluted with buffers containing 0.01% Triton X-100, two washing steps with urea containing buffer with pH 6 (0.01% Triton X-100) were done after washing with urea pH 6 buffer containing 0.5% Triton X-100.

2.17 Protein analysis with a Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)

2.17.1 Tris-glycine based SDS-PAGE

The gel preparation, sample preparation and gel electrophoresis was prepared according to Green and Sambrook (2012) with slight modifications. Modified buffers of SDS gel-loading buffer and tris-glycine electrophoresis buffer were used. The percentage of acrylamide in the gel depended on every experiment. A correct chosen percentage can lead to a better separation of the protein of interest in the gel. The gels were run with 160 V in the electrophoresis apparatus (Biorad®, Mini Protean® 3, #165-3301). In general, 5 µl of protein samples were mixed with 10 µl of SDS gel-loading buffer. Bromophenolblue stained the stacking gel and the protein samples to support visually the loading of the samples. 5 µl protein ladder (several products of Thermo Fisher Scientific®: SM0671, SM0661, #26616) were also loaded in the gel to estimate the molecular weight of the proteins. Gelcode® Blue stain reagent (Thermo Fisher Scientific, #24590) served as Coomassie-blue staining to visualize the separated proteins in the gel and the procedure was done according to manufacture's manual.

2.17.2 Tris-glycerol based SDS-PAGE

Modified separation gel, stacking gel, gel buffer and extra cathode and anode buffer (see 3.1) were used in contrast to previous subchapter. The cathode buffer and anode buffer had to be filled in the appropriate reservoir ensuring no leaking between the both reservoirs occurred. All other aspects were performed according to the previous subchapter. This modified electrophoresis approach was only used in chapter 4.1.2.2 to improve the resolution of protein samples in SDS gels.

2.18 Sample preparation for SDS-PAGE under strongly reducing conditions

This preparation was performed to disrupt protein homo-dimers in SDS-PAA gels. Protein samples were mixed with strongly reducing loading buffer in 1:1 ratio (see 3.1). The samples were heated at 65°C for 10 min and second 95°C incubation followed for 5 min. After the samples cooled down a centrifugation step (13,000 rpm, 2min, RT) and a SDS-PAGE (see 2.17) followed.

2.19 Removal of DNA out of purified protein samples

The DNA had to be removed to exclude any protein-DNA interaction in SDS-gels.

2.19.1 DNase I treatment

The purified protein samples were dialyzed against 1x DNase reaction buffer to be compatible for DNase I treatment. Dialysis was performed according to manufacturer's protocol (scienova™, xpress micro dialyzer 100 MWCO 12-14 kDa, #40784). Additionally a second dialysis step was done against 150 ml stirring 1xDNase reaction buffer for 2 h at RT to ensure a complete dialysis. DNase I (Thermo Fisher Scientific® #EN0521) was added to 100 µl of purified, dialyzed protein solution with final concentration of 0.05 U/µl. An incubation step at 4°C for 30 min and a second step at 37°C for 15 min activated the DNase I reaction. The addition of 2.5 mM EDTA and an incubation at 65°C for 10 min served as deactivation.

2.19.2 Benzonase treatment

The dialysis step was performed similar to the previous subchapter. The dialysis was performed against Benzonase reaction buffer. The dialysis was divided into two steps: the first step was 1 h 45 min long and done on ice followed by substituting the Benzonase reaction buffer and the second step was a 30 min long incubation at RT. Benzonase (Novagen®/EMD Millipore, Benzonase® Nuclease, #70746-4) was added to the dialyzed protein sample with a final concentration of 0.3 U/µl. The incubation step for Benzonase reaction was performed from 30 min to 4 h at 37°C with gentle shaking.

2.20 Detection of recombinant proteins by Western blot analysis

A semi-dry Western blot (alternatively: Immunoblot) was performed to detect the KNL2 variants by His tag and KNL2Cterm specific antibodies. The proteins were separated on a SDS-PAA gel and subsequently blotted on a protein-binding membrane (PVDF or nitrocellulose). Two slightly different approaches were applied.

2.20.1 Western blot according Bjerrum and Schafer-Nielsen (1986)

Procedure was executed according to Bjerrum and Schafer-Nielsen (1986) and manual of the PVDF membrane (GE healthcare biosciences®, Hybond-P PVDF membrane, rpn303F) with slight modifications. A PVDF membrane was used in all experiments as protein binding matrix except in the agarose based EMSA (see 2.22.1), in which a nitrocellulose membrane (GE healthcare, ECL nitrocellulose membrane, rpn203D) was chosen. Electro-blotting was run with 140 mA per blotting stack for 1 h 15 min. However, voltage never exceeded 25 V to avoid damage of the blotting apparatus (BioRad®, Trans-Blot SD, #170-3940). Schleicher & Schuell®´s (GE healthcare) GB005 was chosen as blotting paper. TBS buffer system (see 3.1, Protein transfer buffer) was used in all Western blot experiments. The incubation of the protein-binding membrane in blocking buffer was increased to 2 hours to decrease unspecific background signals. Antibodies were applied according to the related experiments. 'Anti-His tag-HRP' antibodies (monoclonal, mouse, HRP coupled, Miltenyi Biotec®, #130-092-785, 1:5000) was part of direct immunodetection of His tagged recombinant proteins in EMSA experiments (see 4.2.1). Two different pairs of primary and secondary antibodies were used in indirect immunodetection assays. Primary 'anti-His tag' (monoclonal, mouse; Sigma-Aldrich, #H1029, 1:100) and secondary 'anti-mouse-IgG-HRP' (polyclonal, rabbit, HRP-coupled; Sigma-Aldrich, #A9044, 1:500) antibodies were used for the detection of His tagged proteins in Western blots (see 2.20.2 & 4.1.4.2). The primary polyclonal rabbit antibodies against KNL2Cterm ('anti-KNL2Cterm', 1:1000, see 2.27) were detected by the secondary 'anti-rabbit-HRP' (HRP coupled, donkey, GE healthcare, #NA934V, 1:5000) to characterize the KNL2 antibodies and the *KNL2* SALK TDNA insertion line *Arabidopsis* plants (see 2.27 & 4.2.4.4). The incubation with primary antibodies and the antibodies for the direct immunodetection was performed ON at 4°C and the incubation of the secondary antibodies lasted 2h at RT. ECL reaction was done with BioRad® ImmunStar HRP Chemoluminescent kit (#170-5040) according to BioRad®´s manual to trigger the enzymatic, chemoluminescent reaction of the coupled HRP enzyme. GE healthcare´s

Hyperfilm ECL (#28906837) was chosen as chemoluminescent film, which detect the emitted light of the HRP reaction. Exposure time of film to the reaction depends on each experiment. Kodak®-GBX´s developer & replenisher (CAT 190 0984) and fixer & replenisher solution (CAT 190 2485) were used for developing the exposed chemoluminescent film in the dark room. Developing time of the film was 1min, followed by first washing step in water. Then, and fixation was done in 2 min followed by final washing step.

2.20.2 Western blot according to Kyhse-Andersen (1984)

This Western blot alternative protocol can help to overcome some blotting limitations of the first mentioned Western blot approach and it differed in the composition and usage of transfer buffers. The blotting stack was arranged from anode to cathode in following order: 2 blotting papers soaked in western buffer III; 2 blotting papers and the PVDF membrane soaked in western buffer II and the gel from SDS-PAGE and 2 blotting papers soaked in western buffer I. The needed electric current was calculated by 0.8 mA per cm^2 of membrane area. The remaining procedure resembled 2.20.1. This approach was used in the Western blot analysis of the recombinant protein (see 4.1.4.2).

2.21 Protein identification by Mass spectrometry (MS)

Certain bands of KNL2 variants in a SDS-PAA gel were analyzed by MS to verify the identity of these bands. Two spots per protein band were picked (with spotpicker of Biostep®; 1.5mm diameter) out of a Coomassie-stained SDS-PAGE gel (see 2.17). The pieces were separated and subsequently washed with 400 µl of MS wash solution followed by shaking for 30 min at RT. The samples were centrifuged down and supernatants were discarded. Gel pieces were dried by a vacuum concentrator for 15 min. 7.5 µl of MS digestion solution was added to each gel piece and the pieces were incubated for 5 h at 37°C. 1 µl of 1% TFA was added, followed by a centrifugation step. The sample was incubated ON at 5°C. 0.5 µl of the sample was placed on the target plate (Bruker Daltonics, MTP AnchorChip™ 384 800um, # 209514) and dried partially. 1 µl of MS working solution was added and the remaining solution was evaporated. MALDI-TOF analysis was applied to identify the samples.

2.22 Protein-DNA binding analysis by Electrophoretic Mobility Shift Assay (EMSA)

Two different EMSA approaches were chosen to study the DNA binding capability of the purified KNL2 variants (see 2.16.2) and PCR amplified DNA fragment cen180 (see 2.4).

2.22.1 Agarose gel based EMSA

The purified protein samples (see 2.16.2) were dialyzed against EMSA buffer 1 according to manufacturer's protocol (scienova™, xpress micro dialyzer 100 MWCO 12-14 kDa, #40784) and their concentration was determined by Bradford method (see 2.15). The amplified cen180 sequence (see 2.4) was purified (see 2.5) and subsequently eluted in 10 mM tris-HCl pH 7.5. The remaining compounds of EMSA buffer 1 were added to these DNA samples to achieve the same buffer conditions in DNA and protein samples. The DNA concentration was measured by nanodrop (see 2.7). The EMSA was performed according to Palecek et al. (1999) with some modifications. 0.5 µg of DNA and 0.5 µg of protein were used. After the combination of both binding partners the volume of the protein-DNA mixture was not increased to 20 µl to maintain the high DNA and protein concentration. For the formation of protein-DNA complex samples were incubated for 30 min on ice. Then, 5 µl of DNA loading buffer was added to the mixture to enable the loading in 1% agarose gel, which was pre-stained with SYBR® safe (see 2.6). Capillary blotting was performed to transfer samples on the nitrocellulose membrane. The manual of Whatman®/Schleicher&Schuell Turbo™Blotter System (now: GE healthcare) was used with slight modifications to transfer the proteins from the agarose gel to a protein-binding nitrocellulose membrane (GE healthcare®, Hybond ECL nitrocellulose membrane, rpn203D). The initial denaturation step of the agarose gel was not needed to be performed. The agarose gel was shortly rinsed with dH_2O, then, soaked in 20x SSC for 30 min and finally incubated in 6x SSC. The protein-binding membrane was rinsed with dH_2O and incubated for 10 min in 6x SSC. The blotting stack consisted of following parts from the bottom to the top: 20 dry sheets of GB004 (Whatman®/GE healthcare), 4 dry sheets of hybond blotting paper (GE healthcare, rpn6101M; thickness 0.39 mm), one 6x SSC soaked sheet of hybond blotting paper, the prepared membrane, the prepared gel and three 10x SSC soaked sheets of hybond blotting paper. The assembled stack was connected with a 10x SSC soaked ´buffer wick´ to the 10x SSC filled 'buffer tray' of the Turbo™Blotter System. The blotting was performed over 4h 15 min at RT. The membrane was washed in 2x SSC for 5 min after blotting, followed by second washing step in 1x TBS for 5 min. A

Western blot (see 2.20) was performed afterwards to identify blotted proteins on the nitrocellulose membrane.

2.22.2 PAA gel based EMSA

The recommendation of Hellman and Fried (2007) were followed. Purified protein samples were dialyzed similar to previous subchapter. EMSA refolding buffer was used in the first dialysis step and the second step was performed with EMSA buffer 2. PCR of the cen180 fragment was done according to the previous subchapter. The cen180 fragments were eluted with EMSA buffer 2 at the end of the PCR sample purification (see 2.5). Protein and DNA concentration measurements were also performed according to 2.22.1. The volume for 4 pmol of cen180 or KNL2 variants were calculated by Promega®´s Biomath *in silico* tool (http://www.promega.de/resources/tools/biomath-calculators) based on the measured DNA or protein concentration. An native 7% or 8% PAA gel (see 4.2.1) was cast according to Green and Sambrook (2012) with 0.5x TBE. The pre-electrophoresis was performed with 70 V or 10 mA per gel and without the samples in an ice cooled electrophoresis chamber to check the property of the gel. KNL2Cterm and cen180 were mixed according to distinct ratios ranging from 1:1 to 1:5 (see 4.2.1). The DNA competitor poly-deoxyinosinic-deoxycytidylic acid sodium salt (poly dI/dC; Affymetrix, #20539) may be added in concentration ranging from 3-100 µg/µl to verify DNA binding capability of the protein. 1 volume of EMSA additives buffer was added to 10 volumes of DNA-protein mixtures to enable the loading of the samples into the gels and to stabilize the protein-DNA complexes. The electrophoresis was done with 70 V or 10 mA per gel about 1 h with ice cooling. DNA was stained after electrophoresis with SYBR® safe staining (life technologies®, S33102) to visualize the DNA fragments. The SDS-loading of the proteins was performed according to Granger-Schnarr et al. (1988) to enable the proteins to be transferred from the native PAA gel to the PVDF membrane. A Western blot (see 2.20.1) visualized KNL2 variant proteins.

2.23 Protein interaction studies by Bimolecular Fluorescence Complementation (BiFC)

The interaction of two proteins, which are fused with complementary ‘split’ parts of GFP or GFP variants (YFP/EYFP), can be observed in *Nicotiana benthamiana* plants transiently transformed by *Agrobacterium tumefaciens* carrying vector for expression of fusion proteins. Overnight cultures were prepared with 5 ml of YEB medium containing

kanamycin (50 µg/ml) and rifampicillin (50 µg/ml) and this liquid medium was inoculated by glycerol stocks of *Agrobacterium* (strain GV3101) cultures carrying the vectors pSPYCE-35s or pSPYNE-35s (Prof. Dr. Klaus Harter, University of Cologne) with coding inserts of split *YFP* fused to *KNL2*, *KNL2*Cterm and *CenH3*. Incubation of cultures was performed at 28°C with shaking at 180-200 rpm. Overnight cultures were centrifuged down (5000 rpm, 10 min, RT) and the pellets were resuspended in AIM with freshly added acetosyringone (final concentration 0.15 mM). AIM plus acetosyringone were added until the OD600 values of the cell suspensions were diluted to 0.8. The cultures were combined according to following scheme: 2 volumes of the culture with compatible N-terminal part of split YFP, 2 volumes of the culture with compatible C-terminal part of split YFP and 1 volume of HcPro vector containing culture. The co-infiltration of HcPro supported the transient expression of both fusion-constructs. The mixed cultures were incubated at RT for 1 h with moderate shaking at 50 rpm. *Nicotiana benthamiana* leaves were infiltrated with the cultures by needleless syringes (Henke Sass Wolf, Soft Ject®, #5010.200V0). Plants shall be protected for one day against intensive light after infiltration. Infiltrated leaves were analyzed after two or three days under an epifluorescence microscope (Zeiss, Axiophot equipped with a 100x/1.45 α plan-fluar objective and a Sony® DXC-950P color camera). Appropriate excitation and emission filter were used to visualize YFP signals. Some pictures were acquired by applying Differential Interference Contrast (DIC). The lower epidermis was extracted to exert DIC.

2.24 *In silico* co-expression analysis

The program CORNET (https://cornet.psb.ugent.be/main/tool) was chosen for analysing the co-expression of different candidate genes (see 4.2.3.1). Predefined groups of micro array experiments were used and the standard setting was changed by taking the "pair-wise correlations" option to get a comprehensive overview about the candidate genes. All available microarray data were selected.

2.25 Kinase assay

The cDNA *Aurora kinase 3* was inserted into pDEST566 expression vector (Life technologies®). Subsequently recombinant Aurora kinase 3 was expressed in *E. coli* BL21 strain cells and purified against its MBP-tag with amylase resin (New England Biolabs, #E8021). The Aurora kinase 3 and KNL2 variants were dialyzed against 1x kinase assay

buffer (compare 2.26). The kinase assay was composed of following ingredients: 1 µl of ATP, 1 µl of ATP-P32 (Hartmann Analytic, SRP-301), 1 µl of 10x kinase assay buffer, 1 µl of dH_2O, 8 µl of dialyzed KNL2 variant solution and 8 µl of dialyzed Aurora kinase 3 solution. The kinase reaction was carried out for 1.5 h at 30°C. 6 µl of protein loading buffer (Carl Roth®, 4x Roti®-Load 2, K930.1) was added and the samples were incubated at 95°C for 5 min to prepare them for loading on SDS-PAA gels. After the SDS-PAGE the gels were stained with Coomassie (see 2.17). The stained gels were incubated in 3% glycerin for 30 min to prepare them for the subsequent drying procedure. The gels were covered in Cellophane-membranes and subsequently dried (compare Biorad, GelAirDrying system, 165-1771) to increase the phosphorylation signal. The dried gels were analyses with phospho-imager (Fuji, FLA 5000) after one hour exposure.

2.26 Refolding of purified protein samples

Refolding of purified KNL2Cterm (see 2.16.2) was done according to manufacturer's protocol (scienova™, xpress micro dialyzer 100 MWCO 12-14 kDa, # 40784) to remove urea from the buffer and achieve buffer conditions, which were suitable for injection in rabbit. The procedure was separated in three steps. The first dialysis step was performed over 2.5 h against 1300 µl refolding buffer at RT with changing the refolding buffer every 30 min. In the second step, the sample was dialyzed over 3 h 15 min against 1300 µl rabbit injection buffer at 4°C. The buffer was changed frequently every 45 min. In the third step, the sample was dialyzed over night against rabbit injection buffer.

2.27 Production of polyclonal rabbit antibodies

Two rabbits were injected with 500 µl of Freund´s complete or incomplete adjuvant (BD DIFCO; #263810 = complete, #263910 = incomplete adjuvant) and 500 µl of 0.5 – 1 µg/µl purified (see 2.16.2) and refolded (see 2.26) KNL2Cterm protein three times to trigger an immune response. First injection was done with complete adjuvant and the two subsequently following injections were performed with incomplete adjuvant. Second injections were performed 4 weeks after the firsts and one week later the third injections was done. One week after the third injection the blood samples were taken by from ears of the rabbits and about 2 weeks later the rabbits were sacrifice and the antiserum was taken.

2.28 Antibody analysis by Enzyme Linked Immunosorbent Assay (ELISA)

The ELISA was performed to control the immune response of the rabbits against the antigen KNL2Cterm. Nunc® microplates (Th. Geyer, #7650589) were prepared with 7 ml of 1x PBS plus 10 µg/ml of antigen. The microplates were incubated overnight at RT. The supernatant was removed, 200 µl of ELISA blocking buffer per well were added and the plates were incubated for 2 h at RT. 100 µl of diluted ($1:10^4$-$1:10^7$) antisera were added per well followed by an incubation step for 1-1.5 h at RT. The microplates were washed five times with PBS-T. The secondary antibodies (anti-rabbit, IgG, alkaline phosphatase coupled; Sigma-Aldrich, #A3812, 1:2000) was added and the microplates were incubated for 1 h at RT. 15 ml of ELISA detection solution was added to the microplates followed by an incubation step at 37°C for 10 min. The microplates were read out with an ELISA reader (Tecan, Infinite 200).

2.29 Purification of antiserum

The antisera were treated with two different methods, protein A purification and ammonium sulfate precipitation, to remove non-protein parts of the antiserum.

2.29.1 Purification of IgG antibodies using protein A

The protein A-agarose purification was performed with beads of Roche Applied Science (#1 719 408) according to manufactures protocol to isolate IgG isotype antibodies with high affinity. Other antibody isotypes (IgA, IgD, IgE and IgM) were minor targets of the purification. The eluted antibodies were dialyzed (Carl Roth, 27/32, 12-14000 MWCO) against 1xPBS for 1.5 d at 4°C.

2.29.2 Ammonium sulfate precipitation

Following protocol was used for the precipitation of all proteins in the antisera. 1 volume of antiserum and 1 volume of saturated ammonium sulfate solution were mixed and subsequently shaken at 4°C. The solution was centrifuged down at 4000 rpm for 10 min at 4°C. The pellet was dissolved with PBS (usually 10 ml) until the solution got clear and the sample was dialyzed against PBS (see 2.29.1).

2.30 Immunostaining with anti-KNL2Cterm antibodies

Arabidopsis thaliana flower buds or root tips were used as material for the anti-KNL2Cterm antibodies immunostaining treatment to verify these antibodies and analyze the sub-cellular localization of KNL2.

2.30.1 Immunostaining of *Arabidopsis* flower buds

Flower buds were cut from the plants and transferred into MTSB buffer followed by shaking. Fresh immunostaining fixation buffer replaced MTSB buffer. Mild vacuum was applied for 4-5 min terms for total 20 min to let the MTSB buffer infiltrate the plant material. Flower buds were washed 2-3 times with MTSB buffer and 150 µl of immunostaining enzyme mix was added. The enzymatic reaction took place for 20-40 min at 37°C. Another washing step with MTSB buffer was performed. The anthers of the flower buds were separated under a stereomicroscope. The anthers of three flower buds were placed on each slide. Cover slips were laid over the anthers' suspension and the anthers were squashed with moderate strength to enable the removal of cytoplasma. The edges of the cover slip were marked by scratching the slide surface, then, the slides were frozen in liquid nitrogen and the cover slips were removed. The slides were stored shortly in MTSB buffer and subsequently 60 µl of immunostaining blocking buffer were dropped directly onto the sample area. The buffer was covered with Parafilm® (Menasha WI 54952) and incubated for one hour at RT in humidity chamber, then, 40 µl of primary antibodies were added. Primary antibodies were rabbit antiserum immunized by KNL2Cterm (1:500), protein A purified 'anti-KNL2Cterm' (1:500) or 'anti-GFP' dylight® 488 (monoclonal, mouse; Rockland Immunochemicals, 200-341-215, 1:50). Antibodies were diluted in immunostaining blocking buffer. Slides were covered with Parafilm® and stored overnight at 4°C in humidity chamber. Next day slides were washed in MTSB buffer times for 10 min each and subsequently 40 µl of secondary 'anti-rabbit-rhodamine' antibodies (goat, polyclonal, rhodamine coupled; dianova®, 111-295-144, 1:200) was added and samples were incubated for 1 h in humidity chamber. A repetition of the previous washing step followed. To stain chromatin 8 µl of DAPI solution (10 µg/ml DAPI diluted in Vectashield®; Vector Laboratories, #H-1000) was added per slide and the samples were covered by a cover slip. The slides were dried with drying paper and they were analyzed by using an epifluorescence microscope (see 2.23). The microscope was integrated into a 3D digital optical microscope system (Confovis). Structured Illumination Microscopy (SIM) was applied to achieve resolution at 100 nm by using an Elyra

microscope (Zeiss) equipped with a C-Apochromat 63x/1.40 W Korr objective. Lasers were used to excite DAPI (405 nm), dylight® 488 (488 nm) and rhodamine (561 nm).

2.30.2 Immunostaining of *Arabidopsis* roots tips

The procedure for the preparation of root tips was similar to the procedure of flower buds preparation (see 2.30.1) with few modifications. *Arabidopsis thaliana* seedlings were grown on vertically orientated, wet tissue paper over three days and subsequently used for immunostaining. After enzyme treatment the root tips were separated from seedlings and 10 root tips were taken per slide. The squashing was performed more carefully.

3 Materials

3.1 Media

AIM: 10 mM $MgCl_2$, 5mM MES, pH 5.3-5.5

Anode buffer (10x): 2 M tris, pH 8.9

Bacterial lysis buffer: 50 mM tris-HCl, 500 mM NaCl, 15% (v/v) glycerol, pH 8

Benzonase reaction buffer: 50 mM tris-HCl, 1 mM $MgCl_2$, pH 8

Blocking buffer: 1x TBS, 0.1% [v/v] Tween 20, 5% [w/v] dried skimmed milk

Cathode buffer (5x): 500 mM tris, 500 mM tricine, 0.5% [v/v] SDS

DNA loading buffer: 50% [v/v] glycerol, 1 mM EDTA, 0.4% [w/v] bromophenolblue

DNase reaction buffer (1x): 10 mM tris-HCl, 2.5 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.5

ELISA blocking buffer: 1x PBS, 3% BSA, 0.5% Tween 20

ELISA detection buffer: 15 ml 99% diethylamine + 10 g/l $MgCl_2$, pH 9.6; add 1 pill of 4-Nitrophenyl Phosphate Disodium Salt Hexahydrate (Sigma-Aldrich, N2765)

EMSA additives buffer: 10 mM tris-HCl, 50 mM KCl, 1 mM DTT, 0.01% [v/v] Triton X-100, 44% [v/v] glycerol, 110 µg/µl BSA, 0.4% [w/v] bromophenolblue, pH 7.5

EMSA buffer 1: 5 mM tris-HCl, 50 mM KCl, 0.01% [v/v] Triton X-100, pH 7.5

EMSA buffer 2: 10 mM tris-HCl, 50 mM KCl, 1 mM DTT, 0.01% [v/v] Triton X-100, 0.5 M/0.19 M /no urea (see 4.2.1), pH 7.5

EMSA refolding buffer: 1 M urea, 10 mM tris-HCl, 50 mM KCl, 1 mM DTT, 0.01% [v/v] Triton X-100, pH 7.5

GuHCl pH 8 buffer: 6 M Guanidinium chloride, 20 mM NaH_2PO_4, 300 mM NaCl, 5 mM β-mercaptoethanol, 10% [v/v] glycerol, 0.5% [v/v] Triton X-100, pH 8

Immunostaining antibody buffer: 1% BSA in MTSB buffer

Immunostaining blocking buffer: 8 % BSA & 0.1% Triton X-100 in MTSB buffer

Immunostaining enzyme mix: 2.5% pectinase (Sigma-Aldrich® 13789-10G) & 2.5% cellulase (Sigma-Aldrich® C-1184) & 2.5% pektolyase (Sigma-Aldrich® P3026-1G) in MTSB buffer

Immunostaining fixation buffer: 4% paraformaldehyde in MTSB buffer

Kinase assay buffer (1x): 10 mM tris, 50 mM KCl, 2 mM MgCl, 1 mM β-mercapto-ethanol, pH 7.4

MS digestion solution: 10 ng/µl trypsin (Promega, V511A; dissolved in 50mM acetic acid), ≈5 mM ammonium bicarbonate (exactly 4,75 mM), ≈5% acetonitril (exactly 4,75%)

MS wash solution: 10 mM ammonium bicarbonate & 50% [v/v] acetonitrile dissolved in 18 MΩ-H2O

MS working solution: 0.7 mg/ml HCCA dissolved in 90% acetonitril + 1% TFA + phosphate

MTSB buffer: 50 mM PIPES, 5 mM $MgS0_4$, 5 mM EGTA, pH 6.9

LB media: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl

PBS (1x): 8.01 g/l NaCl, 0.2 g/l KCl, 1.78 g/l $Na_2HPO_4 \cdot H_2O$, pH 7.4

PBS-T: 1x PBS, 0.5% Tween 20

PCR mastermix: 14.3 µl dH_2O, 5 µl 5x PCR buffer (Promega 5x Green GoTag® reaction buffer, M791A), 2 µl forward primer (10 µM) (see 3.2), 2 µl reverse primer (10 µM), 0.5 µl dNTP´s (10 mM), 0.2 µl Taq-Polymerase (Promega, GoTag® DNA polymerase, #M830C – for analysis; Takara Bio, PrimeStar®, #R010A – for generation of coding sequence in Gateway®)

Protein transfer buffer (according to Bjerrum and Schafer-Nielsen (1986))**:** 48 mM tris, 39 mM glycine, 20% [v/v] methanol

Rabbit injection buffer: 25 mM urea, 120 mM NaCl, 10 mM Na_2HPO_4, 1 mM DTT, 0.01% Triton X-100, pH 7.5

Refolding buffer: 1 M urea, 280 mM NaCl, 10 mM Na_2HPO_4, 2 mM DTT, 0.01% Triton X-100, pH 7.5

SDS gel-loading buffer: 56 mM Na_2CO_3, 56 mM DTT, 2% [v/v] SDS, 12% [w/v] sucrose

SOC media: 0.5% [w/v] yeast extract, 2% [w/v] tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose (add after autoclaving); Sterilize with 0.2 µm sterile filter

SSC (20x): 3 M NaCl, 300 mM trisodium citrate

Strongly reducing loading buffer: 50 mM glycine, 2% [v/v] SDS, 10% [v/v] glycerol, 5% [v/v] β-mercaptoethanol, 50 mM DTT, pH 2.8

TE (1x): 10 mM tris, 1 mM EDTA, pH 8.0

TBE (1x): 90 mM tris, 90 mM borate acid, 2 mM EDTA

TBS: 50 mM tris, 150 mM NaCl

Tris-glycerol gel buffer: 3 M tris, 0.3% [v/v] SDS, pH 8.45

Tris-glycerol separation gel: 4.88 ml 50% [v/v] acrylamide mix, 8 ml tris-glycerol gel buffer, 8 ml 50% [v/v] glycerol, 3.12 ml dH_2O, 150 µl 10% [v/v] APS, 15 µl TEMED

Tris-glycerol stacking gel: 1 ml 50% [v/v] acrylamide mix, 3 ml tris-glycerol gel buffer, 8 ml dH_2O, tip of bromophenolblue, 160µl 10% [v/v] APS, 16µl TEMED
Tris-glycine electrophoresis buffer (5x): 25 mM tris, 250 mM glycine, 0.1% [v/v] SDS
Urea containing buffer with pH 8/pH 6/pH 5.3/pH 4/pH 2.8: 8 M urea, 20 mM NaH_2PO_4 /(for pH2.8: 20 mM citric acid), 300 mM NaCl, 5 mM β-mercaptoethanol, 10% [v/v] glycerol, 0.5%/0.01% [v/v] Triton X-100*, pH 8/pH 6/pH 5.3/pH 4/pH 2.8
(*TritonX-100 concentration depends on the experiment: 0.5% for analytical western for characterization of protein purification (see 2.19) and 0.01% for EMSA (see 2.22), antibody production (see 2.27) and kinase assay (see 2.25))
Western buffer I: 40 mM aminocaproic acid, 25 mM tris, 20% [v/v] methanol
Western buffer II: 25 mM tris, 20% [v/v] methanol
Western buffer III: 300 mM tris, 20% [v/v] methanol
YEB: 5 g/l beef extract, 1 g/l yeast extract, 5 g/l peptone, 5 g/l sucrose, 0.5 g/l $MgCl_2$

3.2 PCR program and primers

The following PCR schedules refer to the generation of inserts for the entry vector pDONR221 and to the validation of the expression vectors (see 2.1& 4.1). The elongation step of the PCR program was adjusted to the different size of the inserted *KNL2* variants cDNA. The primers for these inserts contain sequences of Gateway® recombinase sites, as well sequences of the recombinant *KNL2* variants. The start codon of the coding sequences and the **end of the cDNA** are marked.

PCR program:
94°C for 5 min (initiation step); 30 cycles of: 94°C for 30 s, 55°C for 30 s, 72°C for 111 s (*KNL2*)/69 s (*KNL2*Nterm)/47 s (*KNL2*Cterm); 4°C (final hold)

Primers:
*KNL2*attB1
5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GAC GGA ACC AAA TCT CGA C-3'
*KNL2*attB2
5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC **TTT** GAT TTT CAA GTT TCT TCG-3'

*KNL2*intattB1

5’-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GAA TTA CTC TGG GAC GAA AG-3’

*KNL2*intattB2

5’-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC **GGG** ATC TAC TAC ATT GTC GTC-3’

Primer pairs for each *KNL2* variant

KNL2 variant	Forward primer	Reverse primer
KNL2	*KNL2*attB1	*KNL2*attB2
*KNL2*Nterm	*KNL2*attB1	*KNL2*intattB1
*KNL2*Cterm	*KNL2*intattB1	*KNL2*attB2

The following PCR schedule is applied in the generation of cen180 products (see 2.4). The cen180 fragments were used in EMSA experiments (see 4.2.1).

PCR program:

94°C for 5 min (initiation step); 30 cycles of: 94°C for 20 s, 60 °C for 30 s, 72°C for 1 min, 4°C (final hold)

Primers:

cen180forward

5’-GGT TAG TGT TTT GGA GTC GAA TAT G-3’

cen180reverse

5’-TTG CTT CTC AAA GAT TTC ATG GT-3’

These primers were used for DNA sequencing (see 2.9 & 4.1.1).

Primers:

pET-T7Pro

5’-TAA TAC GAC TCA CTA TAG GG-3’

pET-T7Ter

5’-GCT AGT TAT TGC GCG G-3’

3.3 Vectors

pDEST17	Life technologies; Carlsbad, CA, USA
pDONR221	Life technologies; Carlsbad, CA, USA
pSPYCE-35s	Walter et al. (2004)
pSPYNE-35s	Walter et al. (2004)

3.4 Organisms

3.4.1 Bacteria

Agrobacterium tumefaciens[*] GV3101	IPK-Gatersleben
Escherichia coli DH5α	IPK-Gatersleben
Escherichia coli BL21	IPK-Gatersleben

3.4.2 Plants

Arabidopsis thaliana Columbia-0	IPK-Gatersleben
Nicotiana benthamiana	IPK-Gatersleben

3.5 Chemicals, enzymes and kits

Affymetrix	(Santa Clara, CA, USA)	DNA competitor
BD DIFCO	(Franklin Lakes, NJ, USA)	Freund´s Adjuvant
Bio-Nobile Oy	(Bio-Nobile Oy, Finland)	magnetic Ni^{2+}-NTA agarose beads
BioRad	(Hercules, CA, USA)	bradford dye, HRP Chemoluminescent kit
Carl Roth	(Karlsruhe, Germany)	several chemicals for media, protein loading buffer
Dianova	(Hamburg, Germany)	anti-rabbit antibodies
EMD millipore	(Billerica, MA, USA)	Benzonase
GE healthcare	(Little Chalfont, UK)	anti-rabbit antibodies
Hartmann Analytic	(Braunschweig, Germany)	radioactive labeled ATP
Kodak	(Rochester, NY, US)	film developer and fixer

[*] alternatively: *Rhizobium radiobacter*

Life technologies	(Carlsbad, CA, USA)	enzymes and vectors for Gateway cloning, agarose, SYBR safe DNA staining
Macherey-Nagel	(Düren, Germany)	Ni^{2+}-NTA agarose beads
Miltenyi Biotec	(B.-Gladbach, Germany)	anti-His tag antibodies
New England Biolabs	(Ipswich, MA, USA)	amylase resin
Promega	(Fitchburg, WI, USA)	Tag-Polymerase, trypsin
QIAGEN	(Venlo, Netherlands)	kits: miniprep, PCR purification, genomic plant DNA extraction; Ni^{2+}-NTA agarose beads
Roche Applied Science	(Penzberg, Germany)	Protein A beads
Rockland Immunochemicals	(Gilbertsville, PA, USA)	anti-GFP antibodies
Serva Electrophoresis	(Heidelberg, Germany)	several chemicals for media
Sigma-Aldrich	(St. Louis, MO, USA)	several chemicals for media, antibodies (anti-His tag, anti-mouse, anti-rabbit), alkaline phosphatase substrate, cell wall digesting enzymes
Takara Bio	(Seta, Japan)	Tag-Polymerase
Thermo Fisher Scientific	(Waltham, MA, USA)	DNA gel marker, protein latter, Coomassie staining, DNase I
Vector Laboratories	(Burlingame, CA, USA)	anti-fade additive in fluorescence microscopy

3.6 Laboratory tools

BioRad	(Hercules, CA, USA)	PCR machine, photometer, electrophoresis apparatus, blotting apparatus
Biostep	(Jahnsdorf, Germany)	gel spot picker
Bruker Daltonics	(Billerica, MA, USA)	MS target plate
Carl Roth	(Karlsruhe, Germany)	dialysis tubing
EMD millipore	(Billerica, MA, USA)	nitrocellulose filter membrane

Eppendorf	(Hamburg, Germany)	photometer
Equi Bio	(Ashford, UK)	electroporation machine (Easy Jet prima)
Fuji	(Tokyo, Japan)	phospho-imager
GE healthcare	(Little Chalfont, UK)	blotting membrane, capillary blotting system, blotting membrane, protein-binding membranes, chemoluminescent film
Henke Sass Wolf	(Tuttlingen, Germany)	needleless syringes
Menesha	(Menasha, WI, USA)	Parafilm
prqlab	(Erlangen, Germany)	nanodrop machine
Retsch	(Haan, Germany)	tissue disruptor
Scienova	(Jena, Germany)	dialysis apparatus
Sonics & Materials	(Newtown, CT, USA)	sonicator
Syngene	(Cambridge, UK)	trans-illuminator
Tecan	(Männedorf, Switzerland)	ELISA reader
Th. Geyer	(Renningen, Germany)	microplates
Zeiss	(Oberkochen, Germany)	epifluorescence microscope, ELYRA microscope

4 Results

4.1 *In vitro* expression of KNL2

4.1.1 Generation and validation of the expression vectors

The KNL2 protein and its N-terminal and C-terminal part (=KNL2 variants) were expressed in *E. coli* to characterize this protein with several approaches (see 1.7). The cDNA of the three *KNL2* variants generated by a RT-PCR reaction using total RNA extract form flower buds of *Arabidopsis thaliana* (Lermontova et al., 2013) were cloned into Gateway compatible vector pDONR221 and subsequently subcloned into pDEST17 vector for protein expression (Supplemental Figure 1, see 2.1).

The presence of the cDNA inserts of expected size inside the expression vector was confirmed by PCR (Figure 1 & Table 1, see 2.3 & 2.4) using flanking primers of the inserts (see 3.2). Subsequently the inserts were validated by sequencing* (Supplemental Figure 2-4, see 2.9). Vectors with correct cDNA insertions in-frame with translation start were selected for protein expression.

4.1.2 Production of recombinant KNL2 variants

4.1.2.1 Analysis of expression of recombinant KNL2 variants in total protein *E. coli* cell extract

KNL2 expression vectors were transformed into *E. coli* BL-21 strain (see 2.2) to express recombinant protein. Protein expression of *KNL2* variants was induced by IPTG in the transformed liquid cultures (see 2.11). Protein extracts were analyzed by SDS-PAGE (see 2.12 & 2.17) with subsequent Coomassie staining of the gel. Additional bands were detected in extracts of IPTG induced cell cultures, in contrast to extracts of uninduced cell cultures (Figure 2). A strong protein band above 40 kDa was observed in the protein extract of bacteria expressing the recombinant KNL2Cterm, additionally a protein band below 70 kDa in extracts of bacteria expressing KNL2Nterm and another above 100 kDa in extracts of bacteria expressing KNL2 extracts were visible.

* In cooperation with IPK Gatersleben

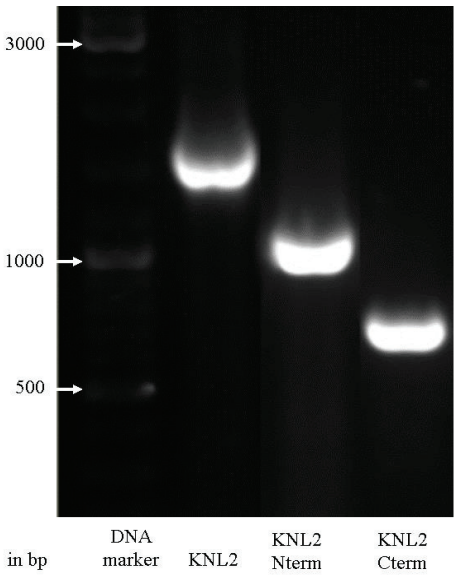


Figure 1: Amplified PCR products with the coding sequences of the three recombinant *KNL2* variants were separated on a 1% agarose gel. A distinct ladder of bands appeared corresponding to the different sizes of cDNA insertions.

Table 1: The expected sizes of the PCR products (Supplemental Figure 1) related to Figure 1 are shown.

KNL2 variant	Expected sizes (in bp)
KNL2	1855
*KNL2*N-term	1147
*KNL2*C-term	769

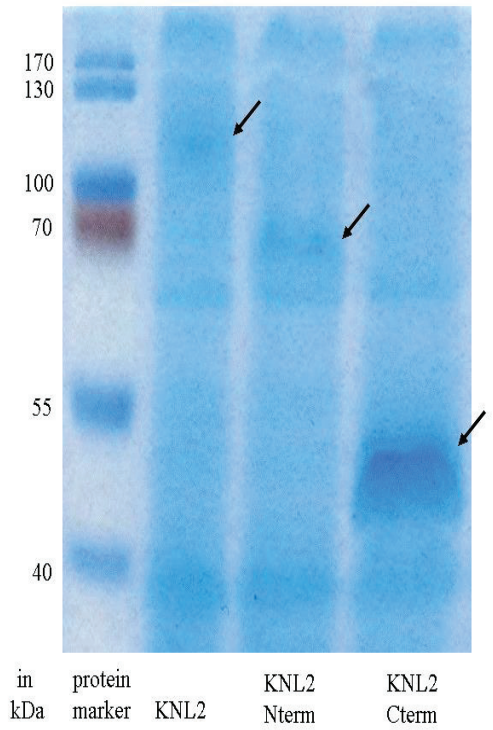


Figure 2: The protein extracts of IPTG induced cell cultures, which were expressing one of the KNL2 variants, were loaded on a 12% SDS-PAA gel. Three distinct bands appear. The KNL2Cterm band show higher intensity.

4.1.2.2 Analysis of solubility of recombinant KNL2 variants

A solubility test was performed to develop a strategy for purification of the recombinant KNL2 protein variants. For this purpose two protocols were used (see 2.13). The outcome

of both approaches was similar. Therefore, the results are shown only for the lab-own protocol. The tris-glycerol based gel electrophoresis (see 2.17.2) was used for this approach. The resolution of the tris-glycerol based gels was comparable to the tris-glycine based ones. The band (see 4.1.2.1) for KNL2 in the total protein extract was visible above 100 kDa and appears also in the insoluble fraction (Figure 3) of this assay. The protein band for KNL2Nterm below 70 kDa was present in the soluble and insoluble fraction. No distinct bands were found in the fraction of *E. coli* extract expressing KNL2Cterm. The band scheme of KNL2Cterm extracts resembled the total extract of non-induced *E. coli* cells (data not shown). We suggest experimental bias was the reason for this phenomenon.

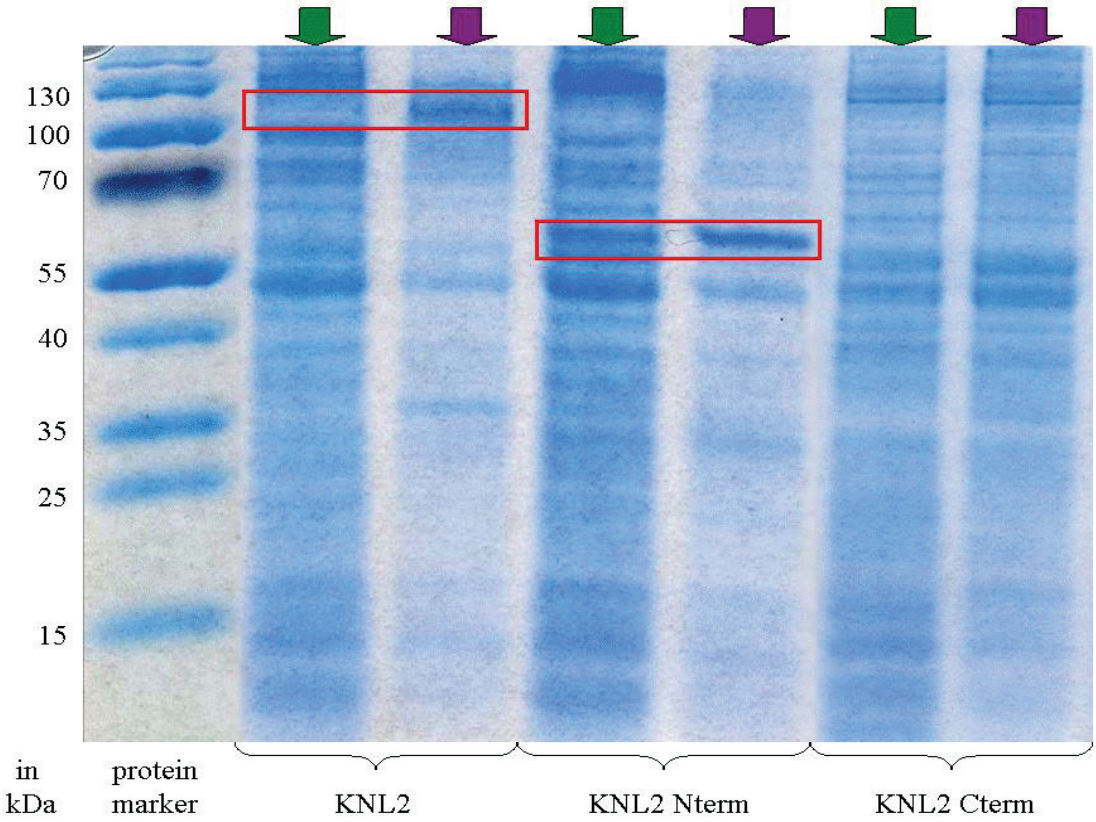


Figure 3: Soluble (green arrowed lanes) and insoluble fractions (violet arrowed lanes) of the solubility assay (see 2.13) were loaded on a 12% SDS-PAA gel. Bands marked by the red frame indicate expressed recombinant KNL2 or KNL2Nterm protein.

4.1.3 Purification of recombinant KNL2 variants

Since, the main part of the KNL2 variants was present in insoluble fraction, a denaturating purification conditions were used. Two different purification protocols were performed (see 2.16). However, only the Ni^{2+}-NTA batch purification allowed to purified sufficient amount of recombinant protein (data of QuickPick™ experiments are not shown). Elution fractions of the batch purification together with total protein extract of the induced cultures

(see 4.1.2.1) were loaded on a PAA gel (see 2.17.2). Already described bands (Figure 2) of the KNL2 variants were present in samples before (blue arrow labeled) and after purification (orange arrow labeled) (Figure 4). Three bands with already observed pattern were identified for the three KNL2 variants: the 110 kDa band for KNL2, the 65 kDa band for KNL2Nterm and the 38 kDa band for KNL2Cterm (Figure 4). There were additional bands below these 'designated main protein bands' in the fractions of the purified protein. These bands were characterized in chapter 4.1.4.2.

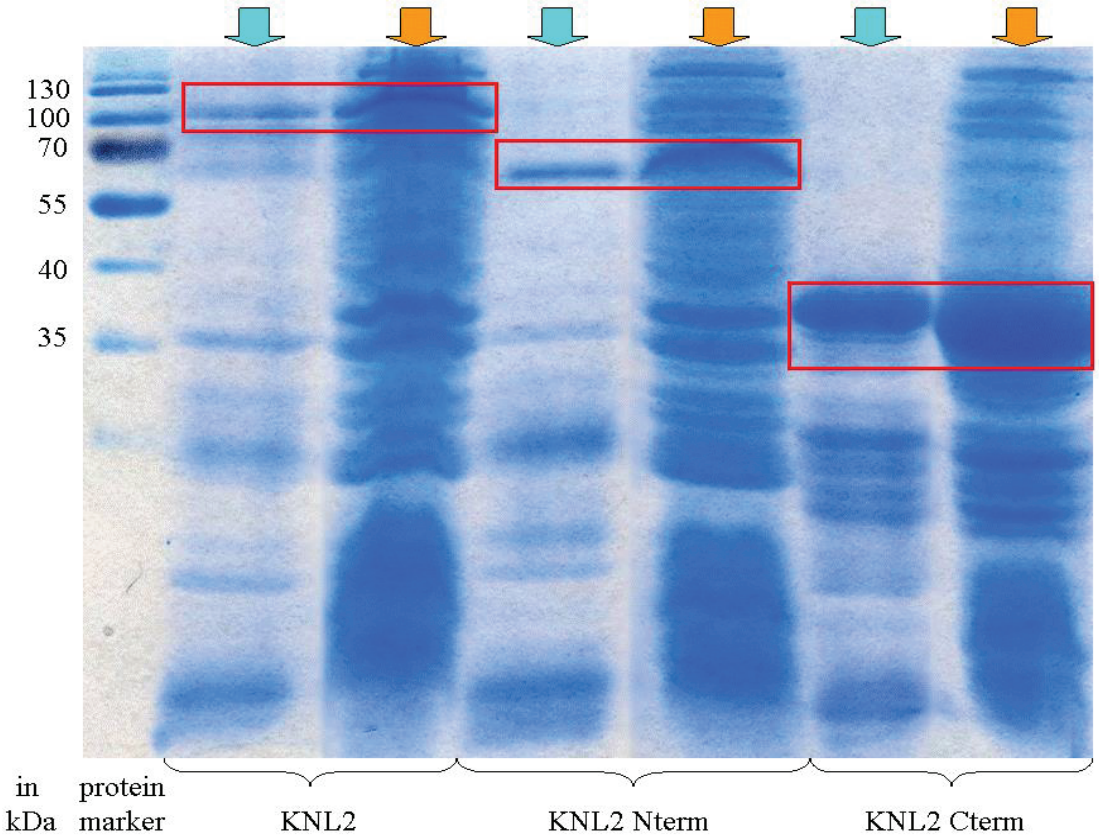


Figure 4: Purified (blue arrowed lanes) and total protein extracts (see 2.12) of induced cultures (orange arrowed lanes) were loaded on a 12% SDS-PAA gel. Designated bands are marked by the red frame indicating expression of recombinant KNL2 variants.

4.1.4 Validation of purified proteins

4.1.4.1 Validation by molecular weight and the molecular weight discrepancy

A correct match between expected and observed molecular weight of the designated bands of the expressed protein are a good hint for the specific expression. Unfortunately that could not be observed for KNL2 and KNL2Nterm (Table 2). KNL2Cterm showed almost a match. However, this observation could not be repeated in all experiments (Figure 2, Figure 4 & Supplemental Figure 5-7). There might be some problems in the execution of

the SDS-PAGE. A discrepancy in molecular weight could be induced by: 1) homo-dimerization, which should leads to a doubled molecular weight, 2) post-translational modification, which should shifts the molecular weight slightly depending on the modification and 3) DNA-binding (see 1.7), which can lead to shift depending on the DNA-protein complex in denaturating PAA gels. Homo-dimerization and DNA-binding of the KNL2 variants were analyzed subsequently. However, no suggestion could be proven experimentally. Therefore, other mechanism or experimental bias may be the underlying reason for molecular weight discrepancy.

Table 2: The expected molecular weights of the recombinant KNL2 variants were compared with the estimated weights of the observed designated bands in the SDS-PAA gel of the purification analysis (Figure 4). The expected molecular weights were calculated from the cDNA sequence (Supplemental Figure 1) by using *in silico* translation (http://web.expasy.org/translate/) and subsequently the translated amino acid sequences served for as input for the calculation of the molecular weight (http://web.expasy.org/compute_pi/). The ratio of observed to expected molecular weights showed high difference for KNL2 and KNL2Nterm to KNL2Cterm.

Protein	Expected molecular weight (in kDa)	Observed molecular weight in (kDa)	Ratio
KNL2	72	≈ 110	153%
KNL2Nterm	46	≈ 65	141%
KNL2Cterm	33	≈ 38	115%

4.1.4.1.1 The molecular weight of KNL2 variants in PAA gels is not changed due to formation homo-dimers

Two different methods were applied to check the hypothesis, that homo-dimerization impacts PAA gel mobility of the KNL2 variants: Strong reducing treatment of the purified KNL2 samples (see 2.18), which should led to disruption of disulfide-bridges based homo-dimers *in vitro*, and a BiFC *in vivo* assay (see 2.23). Within the BiFC assay constructs containing *KNL2* and *KNL2*Cterm fused with two complementary parts of *YFP* separately were transiently expressed in tobacco (*Nicotiana benthamiana*). Restoring of a functional YFP protein was expected only in case of formation of KNL2 or KNL2Cterm homo-dimers. However, no fluorescence could be observed (Supplemental Table 1). This suggests that KNL2 variants do not form homo-dimers. Also, the band pattern did not change through the strong reducing treatment (Supplemental Figure 5) in comparison to chapter 4.1.3. Both assays showed no evidence for homo-dimerization.

4.1.4.1.2 The molecular weight of KNL2 variants in PAA gels is likely not influenced by protein-DNA interaction

The verification of the hypothesis, which assumes a protein-DNA complex was responsible for the molecular weight discrepancy in PAA gels (see 4.1.4.1), was done by the treatment of the protein samples with the enzymes DNase I or Benzonase (see 2.19). These enzymes should digest any DNA in the samples. If there were a shift caused by DNA binding, this effect should disappear after the enzymatic reaction. The band patterns of the KNL2 proteins changed neither by DNase I nor by Benzonase treatment (Supplemental Figure 5 & 6).

4.1.4.2 Anti-His tag antibodies and mass spectrometry verify main bands of SDS gels as KNL2 protein variants and showed truncated products

Two methods were chosen to verify the identity of designated bands of the KNL2 variants (see 4.1.4.1).

A Western blot was applied (see 2.21) to detect all proteins, which carry a His tag, by anti-His tag antibodies. The His tag was placed at the N-terminus of all recombinant *KNL2* variants (Supplemental Figure 1). The elution fractions of the KNL2 variants purification were separated on a 12% PAA gel (identical to Figure 4) and subsequently transferred on a PVDF membrane. The antibodies detected the His tagged proteins on the membrane. The designated bands of KNL2Nterm (65 kDa) and KNL2Cterm (38 kDa) appeared on blot membrane (Figure 5). However, the designated band for KNL2 could not be observed. The percentage of the acrylamide in the PAA gel was lowered to increase the resolution of the protein band of KNL2. Then, a signal for KNL2 was detected at 110 kDa on a blot membrane based on an 8% PAA gel (Figure 6). The additional bands below the designated bands of KNL2 variants were also detected by the anti-His tag Western blot.

The designated bands as well some bands with lower molecular weights were analyzed by mass spectrometry[*] (Supplemental Figure 7, see 2.21). The designated bands of the KNL2 variants were identified as the *Arabidopsis* KNL2 protein (Table 3). Moreover, the identification of the bands of lower molecular weight in the lanes of KNL2 and KNL2Nterm as *Arabidopsis* KNL2 indicates truncation of KNL2 variants at the C-terminal site.

Although the molecular weight discrepancy in PAA gels could not be explained, the purified recombinant KNL2 variants were verified by Western blot and MS.

[*] In cooperation with Dr. Andrea Matros

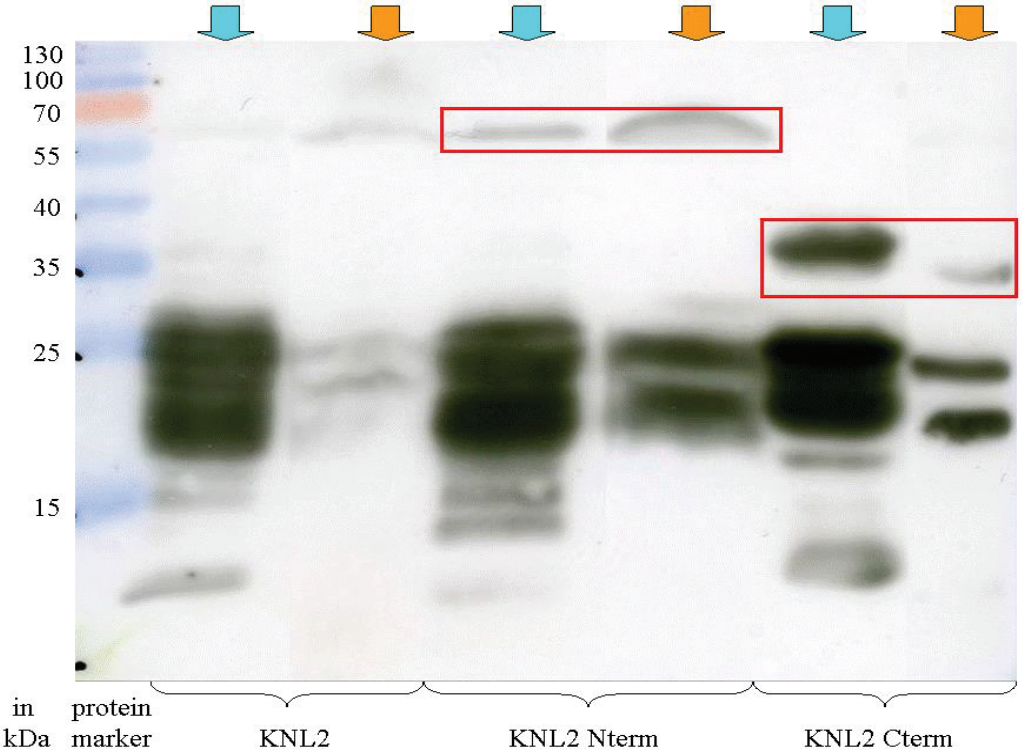


Figure 5: The purified proteins of KNL2 variants (blue arrowed lanes) and the total protein extract (see 2.12) of induced *E. coli* cultures (orange arrowed lanes) were separated on a 12% SDS-PAA gel identical to Figure 4 and subsequently blotted on a PVDF membrane. Primary anti-His tag antibodies and secondary anti-mouse-HRP antibodies were applied on the PVDF membrane. The signals of HRP coupled secondary antibodies were gained on a chemoluminescent film. This film was exposed to the triggered light emitting HRP reaction for 1 min. Bands, which were identical to the designated bands of KNL2Nterm and KNL2Cterm, were highlighted by a red frame.

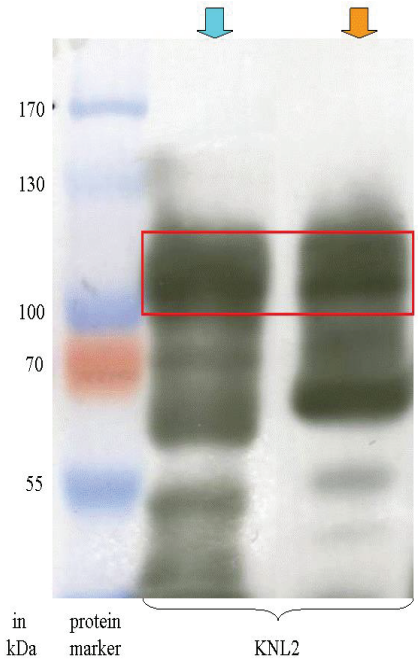


Figure 6: The Western blot was performed according to Figure 5, except the usage of an 8% SDS-PAA gel. The same primary and secondary antibodies were used and the exposure time of the chemoluminescent film to the triggered HRP reaction was 6 h. Signals highlighted by a red frame appeared at the position of designated bands of KNL2 (Figure 4).

Table 3: The picked gel spots of Supplemental Figure 7 were analyzed by MALDI-TOF mass spectrometry. All spots were identified by the older KNL2 database entry gi|7413641 (Supplemental Figure 8) with the exception of the unidentified spot 2b. The MASCOT score above 90 indicate statistical significant hits. Moreover, the number of gained peptides and the sequence coverage of peptides to the whole protein sequence are shown.

Spot/ protein	Protein (NCBI- or BLAST-hit in "plant" or "all entries" database)	Genbank accession	MASCOT Score	Peptides	Sequ. cov. %
1a KNL2	putative protein [*Arabidopsis thaliana*]	gi|7413641	301	5	10
1b KNL2	putative protein [*Arabidopsis thaliana*]	gi|7413641	301	5	10
2a KNL2N	putative protein [*Arabidopsis thaliana*]	gi|7413641	215	2	4
2b KNL2N	Not identified				
3a KNL2C	putative protein [*Arabidopsis thaliana*]	gi|7413641	489	5	10
3b KNL2C	putative protein [*Arabidopsis thaliana*]	gi|7413641	482	6	12
4a KNL2	putative protein [*Arabidopsis thaliana*]	gi|7413641	145	1	2
4b KNL2	putative protein [*Arabidopsis thaliana*]	gi|7413641	346	5	9
5a KNL2N	putative protein [*Arabidopsis thaliana*]	gi|7413641	253	4	6
5b KNL2N	putative protein [*Arabidopsis thaliana*]	gi|7413641	257	4	6

4.2 Characterization of KNL2

The verified, recombinant KNL2 variants were used for the subsequent characterization studies.

4.2.1 Characterization of DNA-binding capability of KNL2

The hypothetical DNA binding of KNL2 protein variants should be analyzed (see 1.7 & 4.1.4.1). Two EMSA approaches were performed (see 2.22): an agarose gel and a native PAA gel based approach. In general, the KNL2 protein variants and the typical *Arabidopsis* centromere repeat cen180 (see 1.3 & 1.4) were mixed. The cen180 repeats were obtained by PCR and subsequently purified (see 2.4 & 2.5). The protein samples were obtained from a urea containing denaturing purification approach (see 2.16.2).

In the agarose gel based EMSA (see 2.22.1) no shift occurred in the lanes of cen180 mixed with KNL2 or KNL2Nterm (Figure 7). However, the band of cen180 mixed with

KNL2Cterm was shifted above 180 bp. The proteins in the agarose gel were transferred on a nitrocellulose blot membrane, where the proteins were detected by anti-His tag antibodies (see 2.20.1 & 2.22.1). Only signals for the KNL2Cterm protein appeared (Figure 8). No signals were detected for KNL2 and KNL2Nterm proteins. A merging of the agarose gel signal into the blot membrane indicated a small shift in DNA and protein bands for the cen180-KNL2Cterm mixture (Supplemental Figure 9).

The EMSA method was changed to a PAA gel based approach (see 2.22.2) to get the signals of KNL2 and KNL2Nterm proteins, because these two proteins might enter PAA gels better than agarose gels. Native PAA gels loaded with KNL2 or KNL2Nterm showed again no shift of the cen180 signal. However, both proteins can be detected by anti-His tag-HRP antibodies (Figure 9 & 10). The cen180 bands were clearly shifted in lanes loaded with a mixture of KNL2Cterm and cen180 sequences (Figure 11). The signal at 180 bp disappeared and shifted near to the slots. A merging of Western blot signals of His tagged proteins into the native PAA gel with SYBR® safe stained DNA indicated overlay of shifted cen180 and KNL2Cterm signals (Figure 12). This indicates a protein-DNA interaction for KNL2Cterm, but not for KNL2Nterm and KNL2.

The urea concentration in the protein-DNA could inhibit protein-DNA interaction in the PAA gel based approach. However, if a shift is observed even in urea containing samples, it must be a protein-DNA interaction (see 5.3.2). On the other hand, proteins could precipitate at low urea concentrations and misfolded or aggregated proteins are supposed to be not able to form protein-DNA complexes. Urea was omitted in the agarose gel based EMSA by dialysis (see 2.22.1). 0.5 M urea was the final concentration in PAA gel based EMSA to ensure solubility. Based on the urea concentration all KNL2 variant samples were partially precipitated in the agarose based EMSA and the samples used in the PAA gel based EMSA were soluble. In an additional approach, the urea concentration was decreased as much as possible in a PAA based approach to find an optimal compromise. KNL2 and KNL2Nterm were still soluble at 0.19 M urea. In general, KNL2Cterm was soluble after removal of urea (see 2.26 & 4.2.4). However, the sample of KNL2Cterm used in this approach precipitated. The protein samples were used, because KNL2Cterm showed in agarose based approach already DNA binding capability, although precipitation was visible in the protein sample indicating soluble proteins were present too. KNL2 variants dialyzed in different urea concentrations (0.5 M, 0.19 M and without urea) were mixed with centromeric cen180 fragments and subsequently a PAA gel based EMSA was performed. The signals of cen180 and KNL2 variants were merged together and the results

indicated no essential difference between the protein-DNA mixtures containing 0.5 M urea and minimum concentrations (Supplemental Figure 10). 0.5 M urea as well as some degree of insolubility might not bias the shift of the cen180 band. Two minor unexpected findings were observed: a decreased signal of cen180 at 180 bp in the lane of the cen180-KNL2Nterm mixtures containing less urea, and a larger shape of the KNL2Cterm protein signal in the lane containing no urea.

A competition assay approach with poly dI/dC was performed to verify the specificity of the KNL2Cterm induced shift of cen180 bands. The KNL2Cterm protein sample without urea was used. No shift at all was observed (Figure 13). Experimental bias might happen.

The EMSA experiments show no DNA interaction of cen180 repeats and KNL2 and KNL2Nterm at all. Shifts for KNL2Cterm containing mixture could be shown indicating DNA binding capability. However, a specific DNA binding could not be proven.

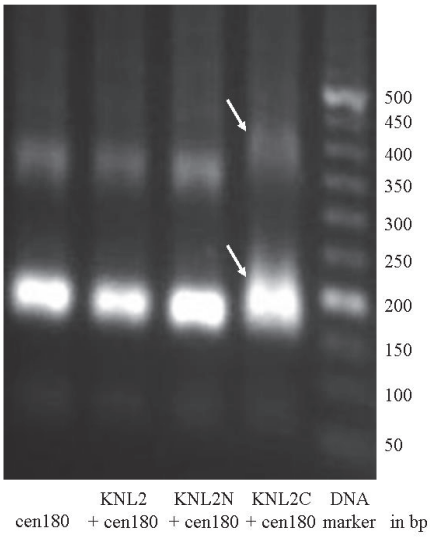


Figure 7: A 1% agarose gel based EMSA was performed to check the interaction of the centromeric sequence cen180 with the KNL2 variants. No shift can be observed for the KNL2 and KNL2Nterm (KNL2N) mixture in comparison to the 180 bp big cen180 band, whereas the mixture of KNL2Cterm (KNL2C) and cen180 showed a small shift (marked by arrows).

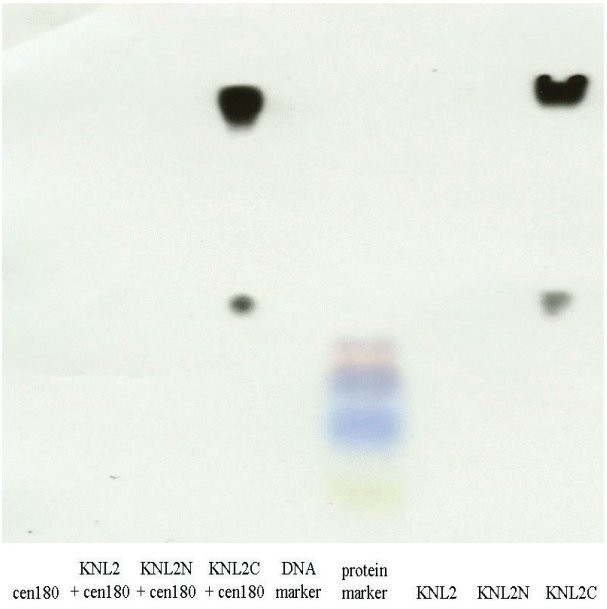


Figure 8: The 1% agarose gel (Figure 7) of the EMSA experiment was blotted on a nitrocellulose membrane by capillary blotting. Subsequently the recombinant KNL2 variants were bound by anti-His tag-HRP antibodies and the emitted light caused by the triggered HRP reaction was detected on a chemoluminescent film after 5 s exposure. Only signals were detected in lanes containing KNL2Cterm.

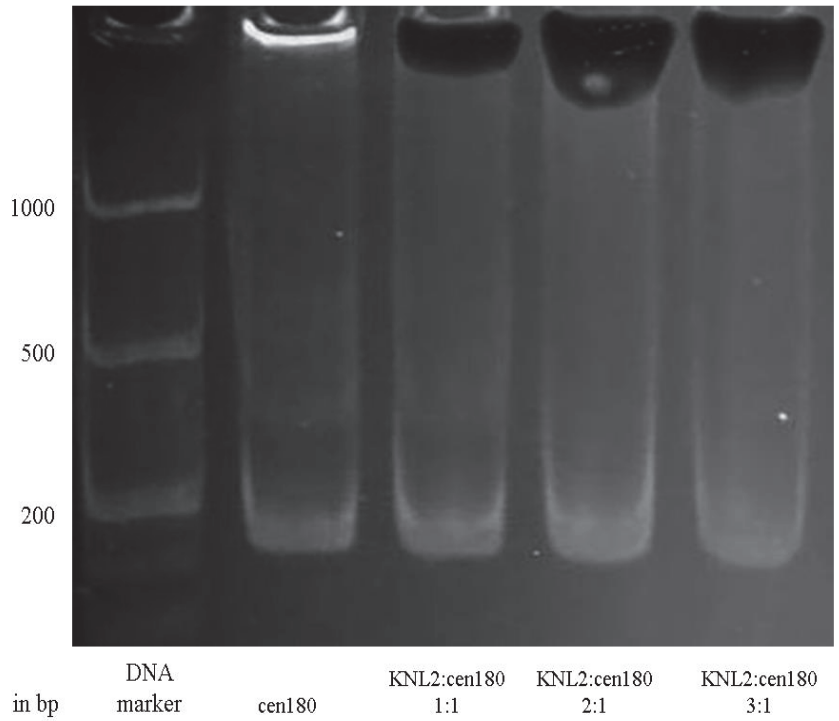


Figure 9: The protein KNL2 and centromeric sequence cen180 were mixed according to defined molecular ratios and loaded on an 8% native PAA gel. The band of cen180 did not shift through rising amounts of KNL2. The proteins were denaturated with SDS and transfer on a PVDF membrane. Anti-His tag-HRP antibodies was used to detect KNL2 proteins. After the HRP reaction the chemoluminescent film was exposed for 10 s. Signals of the chemoluminescent film were merged into the native PAA gel and signals for KNL2 were visible at the slots of each KNL2 containing lane.

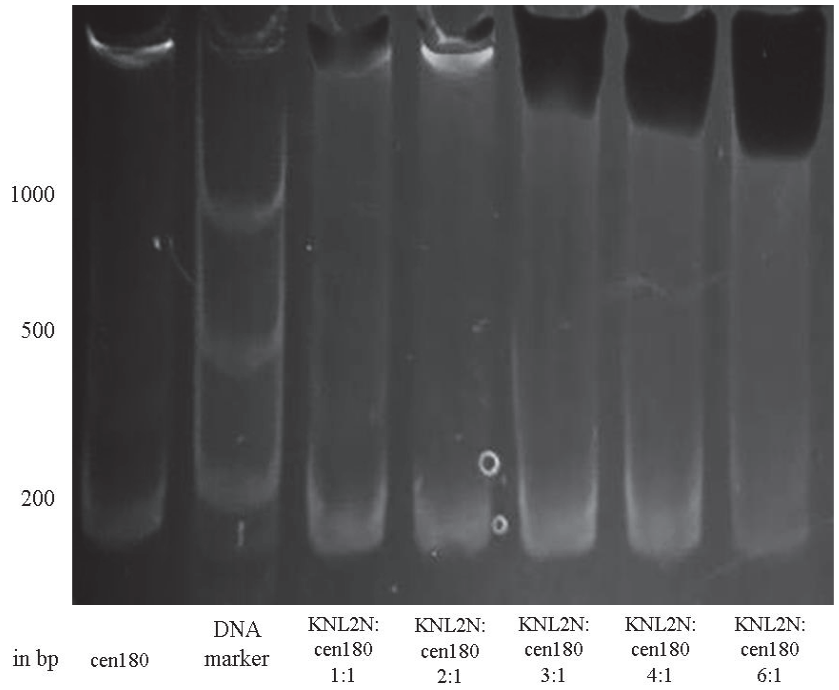


Figure 10: The protein KNL2Nterm (KNL2N) and the cen180 sequence were mixed according to defined molecular ratios and loaded on an 8% native PAA gel. There is no shift for the 180 bp big band of cen180 observed due to KNL2Nterm. KNL2-Nterm were loaded with SDS and subsequently transferred on a PVDF membrane. A Western blot with anti-His tag-HRP antibodies was performed. The HRP reaction was detected on a chemoluminescent film for 10 s. The signals of KNL2Nterm can be observed at the slots of each lane, where KNL2Nterm was loaded.

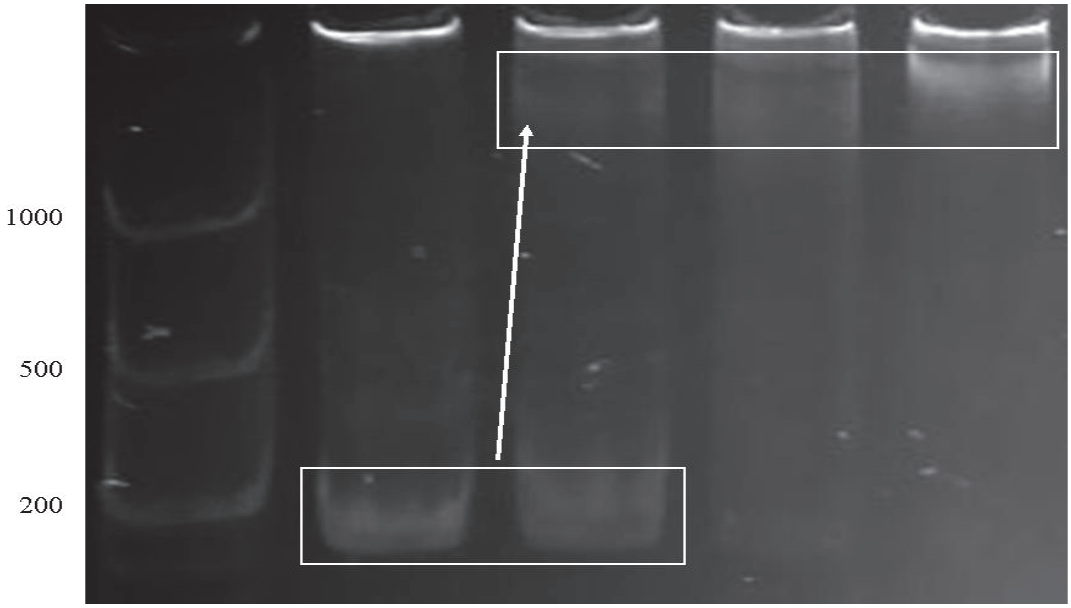


Figure 11: KNL2Cterm protein (KNL2C) and the centromeric DNA sequence cen180 were mixed according to defined molecular ratios and loaded on an 8% native PAA gel. There was a shift of the cen180 band at 180 bp to a position near to the slots. The shift got clearer with higher molecular ratio of KNL2Cterm.

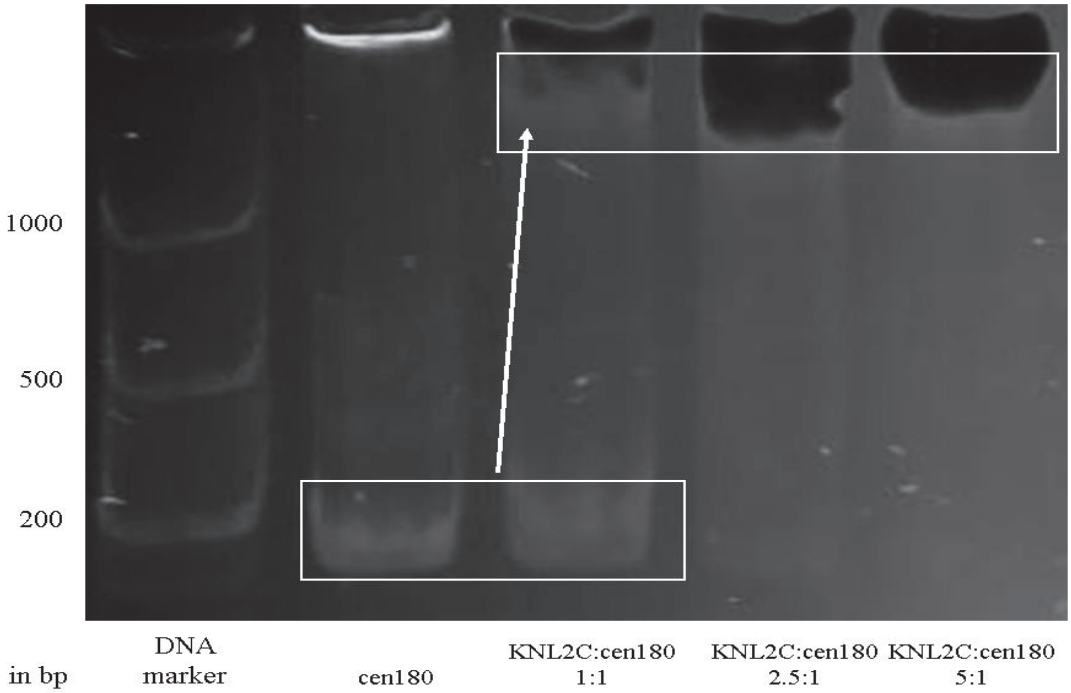


Figure 12: The KNL2Cterm (KNL2C) proteins, which were included in the EMSA experiments of Figure 11, were loaded with SDS and subsequently detected by Western blot (see 2.20) with anti-His tag-HRP antibodies. The signals of the HRP reaction were detected on a chemoluminescent film with an exposure time of 10 s and were merged in Figure 11. There was an overlay of signals of the shifted centromeric cen180 band and the signals of the antibodies detected KNL2Cterm protein.

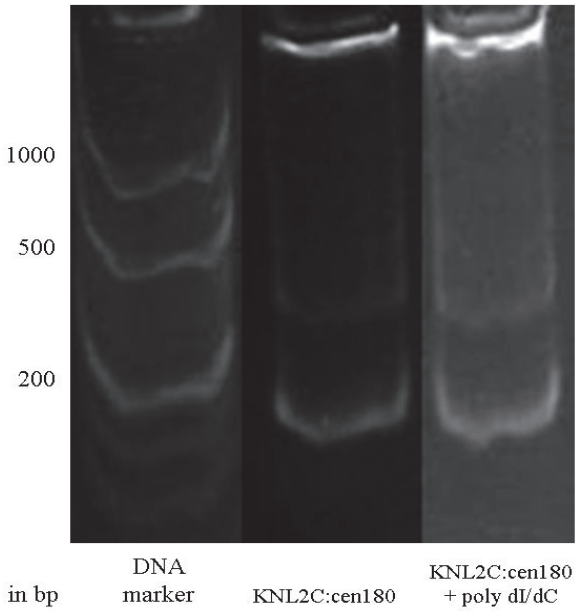


Figure 13: A PAA gel based EMSA experiment (see 2.22.2) was performed with the centromeric cen180 sequence and the KNL2Cterm (KNL2C) protein. The protein-DNA mixture and mixture with the DNA binding competitor poly dI/dC were loaded on a native 7% PAA gel. The DNA was stained by SYBR® Safe. No shift of the 180bp band of cen180 was observed.

4.2.2 Characterization of the interaction of KNL2 and CenH3

Human KNL2 does not interact with CENPA directly (see 1.6.1). The interaction of the two corresponding *Arabidopsis* homologues was checked with a BiFC assay (see 2.23) to analyze their interaction *in planta*. Fluorescence signals were obtained by KNL2- and KNL2Cterm-split YFP fusion proteins co-expressed with CenH3-split YFP fusion in tobacco. Also, CenH3 fusion controls without KNL2 variants showed signals (Supplemental Figure 11 & Supplemental Table 1). Therefore, neither the existence of a binding nor the non-existence could be proved. The experiments has to be repeated and it is recommended to include CENPC homologue (see 1.6.1) or other protein-protein interaction studies like co-immunoprecipitation or yeast/bacterial two hybrid selection systems might be applied (Golemis (Ed.), 2002).

4.2.3 Characterization of possible phosphorylation of KNL2

The suggested phosphorylation of KNL2 shall be analyzed in this chapter (see 1.6.2).

4.2.3.1 *In silico* analysis of co-expression of *Arabidopsis Aurora kinases* and *KNL2*

In silico analysis (see 2.24) was performed to characterize the expression of *KNL2* and certain candidate kinases as well as *CenH3*. The results of the analysis using the software

CORNET and a set of microarrays predefined by the program (in total 425 experiments) showed a significant correlation of expression of selected genes (correlation coefficient above 0.5) (Figure 14).

Aurora kinase 1, Haspin kinase Aurora kinase 3 seems to be reliable candidates for KNL2 kinases.

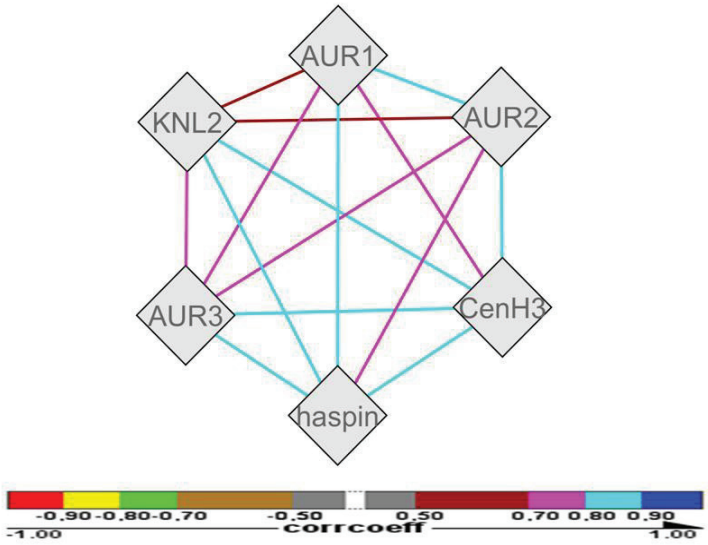


Figure 14: The *in silico* co-expression analysis (see 2.24) compared following *Arabidopsis* candidates: *KNL2*, all three Aurora kinases (AUR1-3 = *Aurora kinase 1-3*), *CenH3* and *Haspin kinase* as well. The color of the lines between the candidates represents the Pearson correlation coefficient.

4.2.3.2 Aurora kinase 3 phosphorylates KNL2 *in vitro*

Aurora kinase 3 was chosen for the kinase assay* (see 5.3.3.2). Only one kinase was used first, because of the limited time. The kinase assay reaction included two denaturing PAA gels, one with the three KNL2 variants and Aurora kinase 3 (Figure 15) and another one without kinase (Supplemental Figure 12). All three KNL2 variants and Aurora kinase 3 bands showed phosphorylation signals in the kinase reaction gel (Figure 16). Minor signals were visible in the control reaction without Aurora kinase 3 for KNL2Nterm and KNL2Cterm (Supplemental Figure 13). However, the signal strength was not comparable to the assay with Aurora kinase 3.

These results showed a phosphorylation at N- and C-terminal sites of KNL2 by Aurora kinase 3 and putative Aurora kinase phosphorylation sites were found within N- and C-terminal parts by amino acid sequence analysis (Supplemental Figure 14).

* In cooperation with Dr. Dmitri Demidov and Eva Tomaštiková

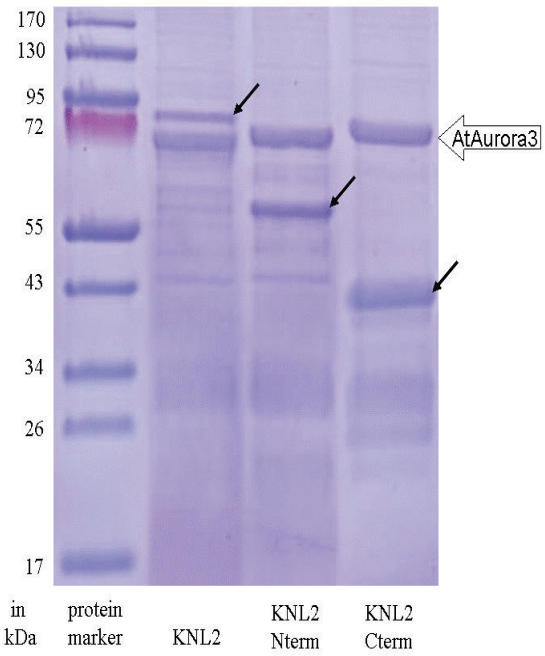


Figure 15: The three KNL2 variants, Aurora kinase 3 and other ingredients of the kinase assay (see 2.25) were loaded on a 10% SDS-PAA gel, separated by SDS-PAGE (see 2.17.1) and subsequently stained by Coomassie. The designated bands of KNL2 as well Aurora kinase 3 bands were marked by arrows. The KNL2 variants were loaded unequally.

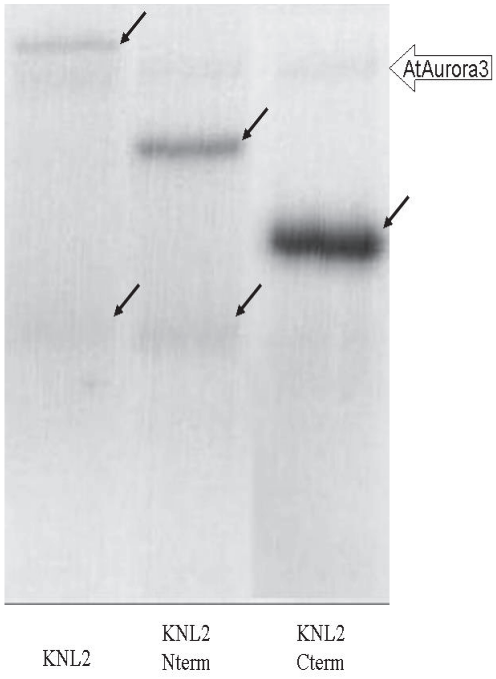


Figure 16: The SDS-PAA gel of Figure 15 was analyzed by a phosphor-imager (see 2.25). Black bands indicate phosphorylation activity. The designated bands of the KNL2 variants, truncated products and Aurora kinase 3 were marked.

4.2.4 Analysis of specificity and application of anti-KNL2Cterm antibodies related assays

Many methods of proteins characterization like immunostaining and Western blot rely on antibodies. Therefore, KNL2 specific antibodies were produced. KNL2Cterm was chosen as antigen, because this variant delivered the highest amount of purified protein (Figure 4) and KNL2Cterm were able to be refolded in a dialysis buffer without urea (see 2.26), while maintaining its solubility.

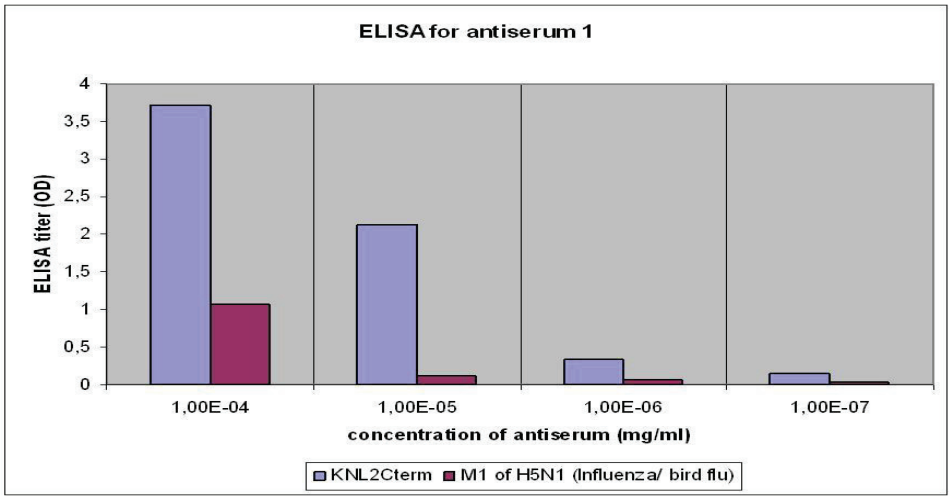


Figure 17: The gained antiserum of rabbit 1 was added in different concentrations to two fixed antigens: KNL2Cterm and His tagged M1 protein of the *Influenza* virus. The bound anti-KNL2Cterm antibodies of the antiserum were detected by alkaline phosphatase coupled anti-rabbit antibodies. The triggered alkaline phosphatase reaction was measured by an ELISA reader. The ELISA titer characterized the binding of the antiserum to the antigen.

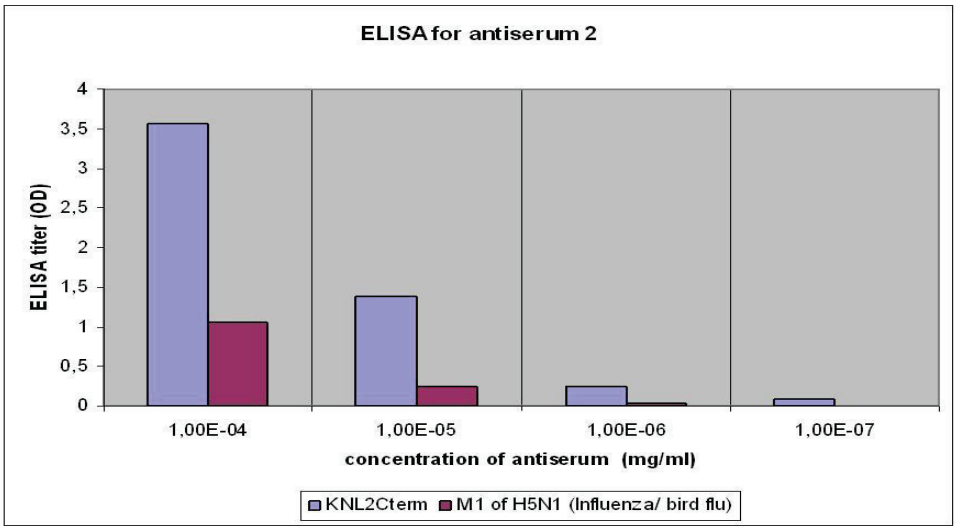


Figure 18: An ELISA identical to Figure 17 was performed to characterize the antiserum gained from rabbit 2.

4.2.4.1 ELISA and Western blot experiments confirmed sensitive immunoreaction of anti-KNL2Cterm antibodies

KNL2Cterm protein samples were injected* in two rabbits (named '1' & '2') according to 2.27. Subsequently ELISA* was performed (see 2.28) to analyze the gained antisera. The ELISA results (Figure 17 & 18) show high ELISA titer of both antisera against KNL2Cterm. In contrast, the antisera gave less response to the control protein, which had just the His tag in common with KNL2Cterm. The antisera were purified or precipitated† according to chapter 2.29.

Then, a Western blot was performed to quantify the reaction between purified or precipitated antibodies to the antigen KNL2Cterm. The results of the anti-KNL2Cterm antibodies 1 (of rabbit 1) are shown in Figure 19 and the antibodies 2 (of rabbit 2) gave similar Western blot signals (data not shown). Even 150 pg of the KNL2Cterm protein can be detected by Western blot using anti-KNL2Cterm antibodies 1. The expected designated band of KNL2Cterm appeared below 38 kDa. Moreover, bands above this molecular weight at 70 kDa were detected, indicating a dimer formation. Furthermore, the expected band for KNL2 at 110 kDa was visible, whereas no signals for KNL2Nterm appeared.

The antibodies showed a good and sensitive response against the recombinant antigen, which indicated a working immune reaction against KNL2Cterm of both rabbits.

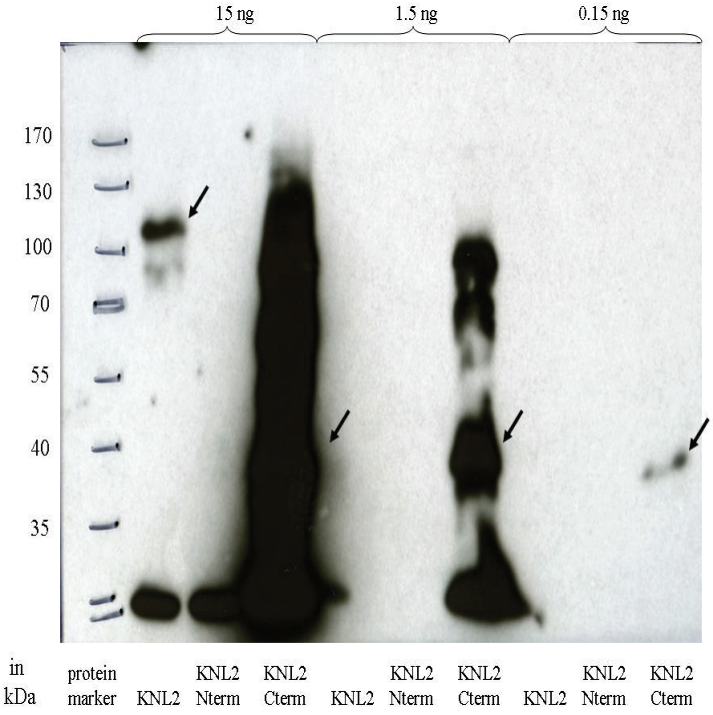


Figure 19: Different amounts of the KNL2 variants were separated on an 8% SDS gel. Subsequently the gel was blotted on a PVDF membrane. Anti-KNL2Cterm antibodies 1 (Figure 17) and anti-rabbit-HRP was applied on the membrane. The triggered HRP reaction was exposed to a chemoluminescent film for 30 min. Arrows indicate the designated bands of KNL2 (110 kDa) and KNL2Cterm (38 kDa).

* In cooperation with Ulrike Gresch
† In cooperation with Ulrike Gresch and Isolde Tillack

4.2.4.2 Localization of KNL2 at chromocenter and in nucleoplasm of meristematic nuclei

The anti-KNL2Cterm antibodies were used to perform immunostaining (see 2.30) on meristematic cells of *Arabidopsis* as well as in cells of the *KNL2* SALK mutant. The *KNL2* SALK mutant is T-DNA-insertion line with a knock-out of *KNL2* expression by a kilobase-big insertion in exon 4 of the *KNL2* gene (Supplemental Figure 15). The immunostaining with the antibodies 1 & 2 on root tips and flower buds of *Arabidopsis* showed similar results. Specific, chromocentric signals and signals in nucleoplasmic background were seen on meristematic cells of *Arabidopsis*, whereas cells of *KNL2* SALK mutants were stained at nucleoplasm only (Figure 20). The chromocentric signals were characterized by centromeric regions surrounded by heterochromatic, pericentromeric regions (see 4.2.4.3). No chromocentric signals of KNL2 antibodies were seen in M phase. KNL2 localize at chromocenteres only in interphase.

4.2.4.3 Co-localization of KNL2Cterm and CenH3 in meristematic nuclei

The anti-KNL2Cterm antibodies 1 and anti-GFP antibodies were used on root tips of *EYFP-CenH3* stably transformed *Arabidopsis thaliana* plants (see 2.30.2). The signals of CenH3 and KNL2Cterm co-localized (Figure 21). Also, this co-localization could be observed at typical *Arabidopsis* chromocenters. SIM technology[*] was applied to get a better resolution (see 2.30.1). The co-localized signals of KNL2 and CenH3 are embedded into the heterochromatic pericentromere (Figure 22).

4.2.4.4 Analysis of *Arabidopsis* ecotype Col-0 and *KNL2* SALK protein extracts

The whole cell protein extract of *Arabidopsis thaliana* ecotype Columbia-0 seedlings and *KNL2* SALK mutant seedlings (see 2.14) were used for Western blot with anti-KNL2Cterm antibodies. The result of the Western blot with anti-KNL2Cterm antibodies 2 is shown in Figure 23. A band above 70 kDa appeared in the lanes of the wild type, however, not in the mutant. The expected molecular weight of KNL2 is 66 kDa. Some other bands with different molecular weight were visible on the film. They were present in the lanes of wild type and mutant samples.

Thus, the preliminary data characterized the gained anti-KNL2Cterm antibodies as specific and sensitive.

[*] In cooperation with Dr. Veit Schubert

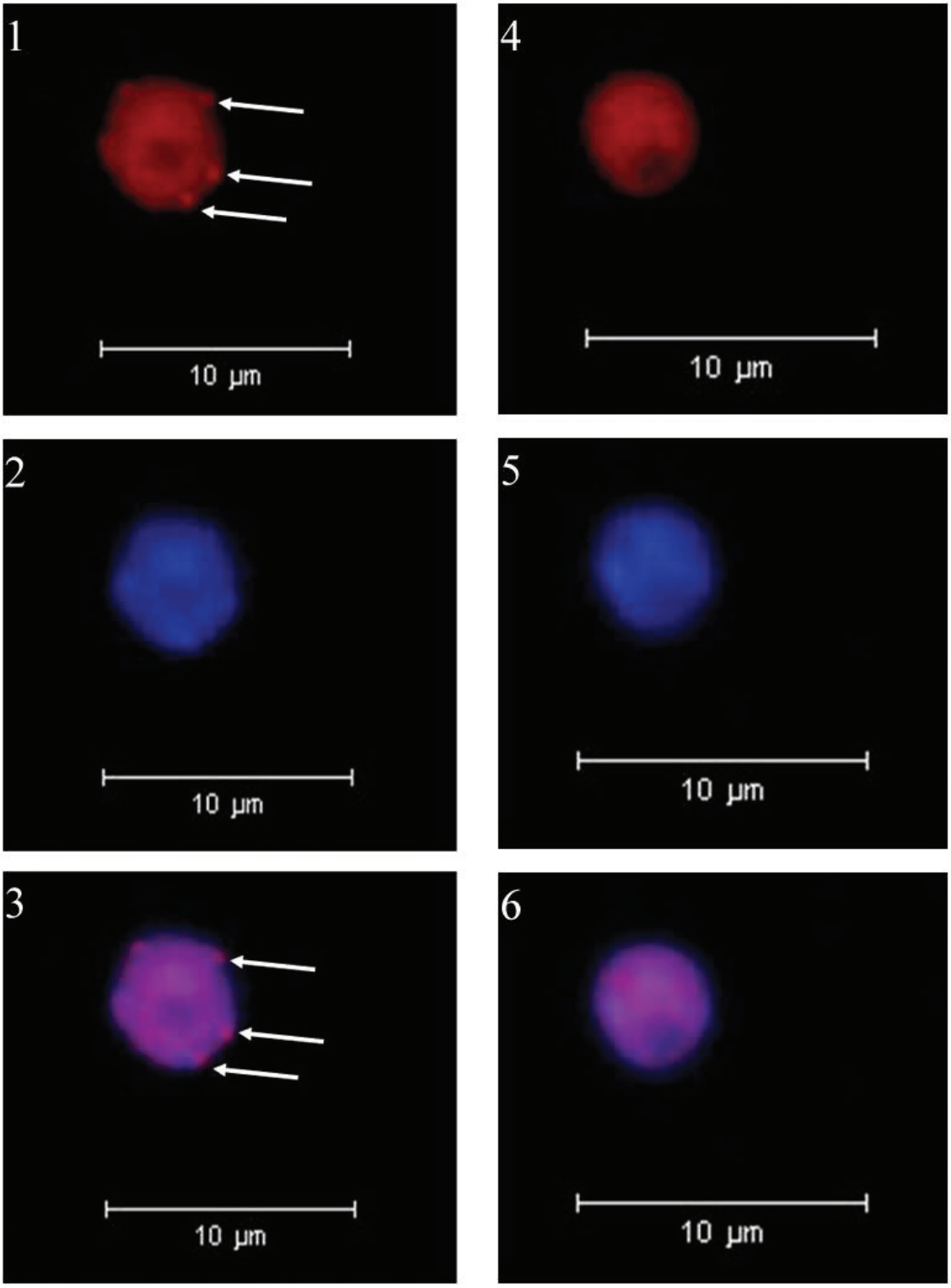


Figure 20: Immunostaining (see 2.30) of flower bud cells of *Arabidopsis* **(1-3)** and flower bud cells of *KNL2* SALK *Arabidopsis* mutants **(4-6)** was performed using anti-KNL2Cterm antiserum 2 and anti-rabbit-rhodamine antibodies. **(1/3)** Rhodamine signals of the secondary antibodies were detected. **(2/5)** DAPI signals of the chromatin were shown. **(3/6)** Rhodamine and DAPI signals were merged. White arrows indicate chromocenters.

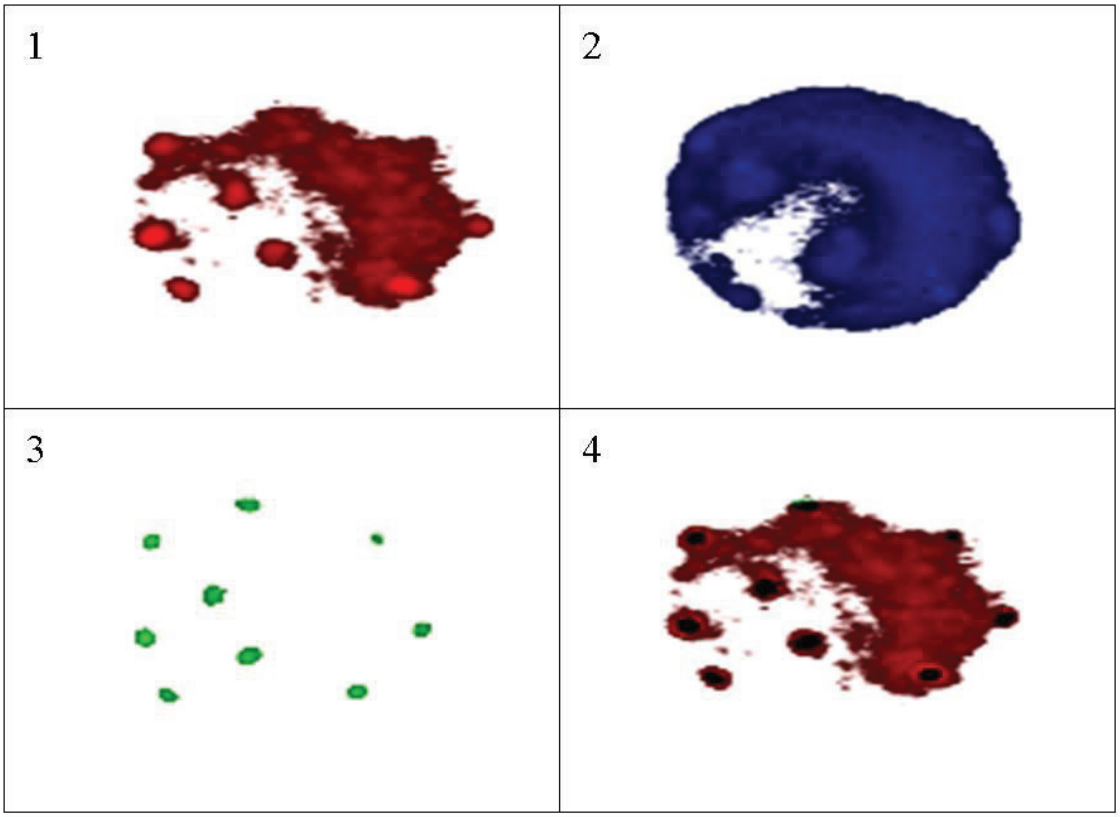


Figure 21: Immunostaining of root tips of *Arabidopsis* plants stably expressing *EYFP-CenH3* fusion constructs with anti-GFP antibodies and protein A purified anti-KNL2Cterm antibodies 1 (Figure 17, see 2.29.1). 3D microscopy technique was applied for analysis of fluorescence (see 2.30.1). **(1)** Rhodamine signals for KNL2. **(2)** DAPI signals for chromatin. **(3)** Dylight® 488 signals for CenH3-EYFP fusion protein. **(4)** Merge of rhodamine and dylight® 488 signals.

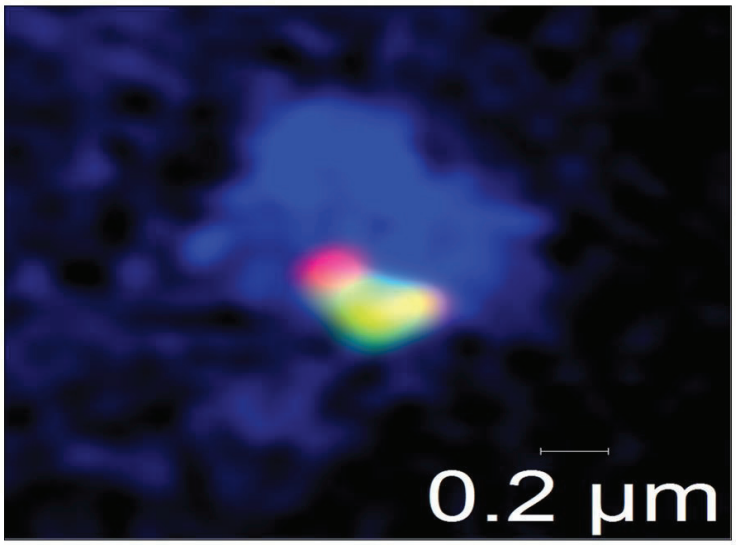


Figure 22: The Structured Illumination Microscopy was applied on immunostaining samples of Figure 21. The red color shows the position of KNL2. CenH3 is represented by the green color. Yellow indicate the co-localization of CenH3 and KNL2. DAPI stain chromatin. Condensed blue staining is condensed, heterochromatic, pericentromeric DNA.

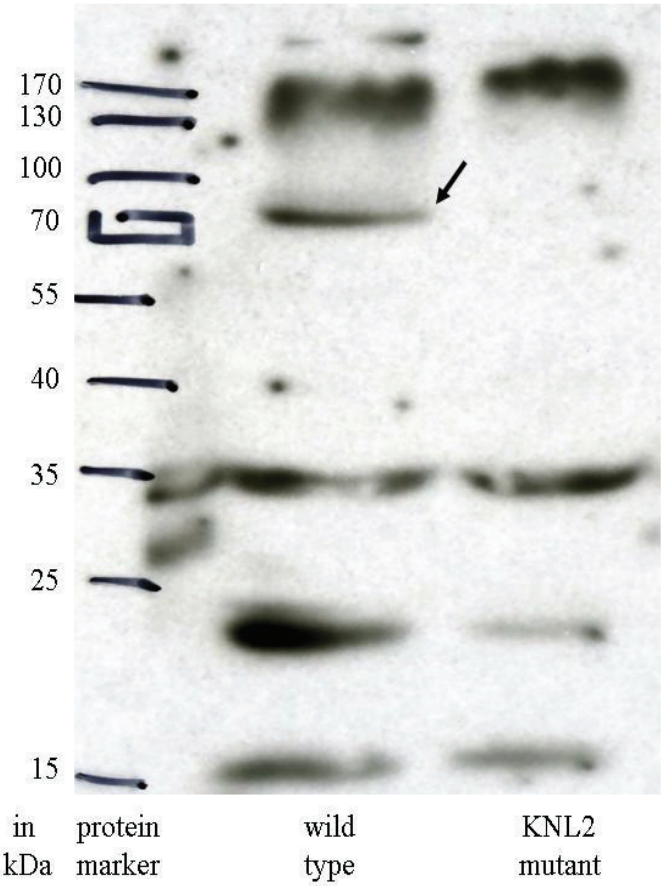


Figure 23: 20 µg of protein samples isolated from seedlings of *Arabidopsis thaliana* ecotype Col-0 (wild type) and *KNL2* SALK mutant were separated on a 10% SDS-PAA gel. The protein samples were transferred from the gel to a PVDF membrane. Ammonium sulfate precipitated anti-KNL2Cterm antibodies 2 were applied (Figure 18, see 2.29.2). Subsequently the secondary anti-rabbit-HRP antibodies were used. The triggered HRP reaction was captured by a chemoluminescent film for 5 min. The arrow marks bands for KNL2.

5 Discussion

5.1 Main findings of the study

The recombinant KNL2 proteins were successfully expressed and purified. The C-terminal part of KNL2 showed indications for a DNA binding capability, but not the N-terminal part and the full length fragment. KNL2 was phosphorylated on multiple sites by Aurora kinase 3. CenH3 and KNL2 co-localized at chromocenters of meristematic *Arabidopsis thaliana* interphase nuclei.

5.2 KNL2 protein variants are purified, however, some non-expected observations arise

The specificity of purified KNL2 protein variants were validated by mass spectrometry and Western blot analysis (see 4.1.4.2). Therefore, all confirmed KNL2 variants were used for further characterization (see 5.2.2).

Nevertheless, two experimental observations were not expected: the KNL2 protein variants showed not the expected molecular weights in SDS-PAA gels (see 5.2.1) and some products of smaller molecular weight appeared additionally to the designated main protein bands (see 5.2.2).

5.2.1 Posttranslational modification or experimental conditions influence the size discrepancy of KNL2 variants in PAA gels

The KNL2 variants showed not expected molecular weight in SDS-PAA gels. We supposed different possibilities for this size differences (see 4.1.4.1): post-translational modification, homo-dimerization and DNA binding.

Certain posttranslational modifications like phosphorylation and glycosylation can lead to an impaired SDS of the proteins (Weber et al., 1972; Green and Sambrook, 2012). Phosphorylation and glycosylation are post-translational modifications in eukaryotes and prokaryotes (Dell et al., 2010). The recombinant, over-expressed eukaryotic KNL2 proteins may be modified in *E. coli* by modifications like the non-enzymatic gluconolation (Stryer, 1995; Aon et al., 2008). Moreover, recombinant KNL2 proteins have a lot of aspartic acid, histidine, serine and threonine residues, which might be phosphorylated by bacterial kinases (Supplemental Figure 14) (Pereira et al., 2011).

Proteins with higher charge might migrated differently in SDS-PAA gels than proteins of the same size with neutral net charge (Weber et al., 1972; Green and Sambrook, 2012). KNL2 and KNL2Cterm have an increased isoelectric point (pI) (Supplemental Table 2). However, these values are not comparable to the pI of CenH3 and other histones, which show atypical electrophoretic mobility caused by their high positive net charge (Weber et al., 1972).

Our data did not support the hypothesis, that protein-DNA interaction lead to the size discrepancy of KNL2 protein variants in SDS-PAA gels (see 4.1.4.1.2). Usually protein-DNA complexes are relative labile even in native PAA gels and rapid dissociation may occur during electrophoresis (Hellman and Fried, 2007; Sidorova et al., 2010). Therefore, presence of stabile protein-DNA complexes, especially in denaturating SDS-PAA gels, may be regarded as unlikely. However, some investigators report about protein-DNA complexes in denaturating gels (Naryzhnyi, 1992). Nevertheless, this phenomenon was not the proven in this study.

Several investigators found stable dimers of other proteins in SDS-PAA gels (Gherardi and Coffer, 1991; Isbert et al., 2012; Hauptmann et al., 2013). However, our experimental data do not show a homo-dimerization of KNL2 variants (see 4.1.4.1.1).

We were unable to verify our suggestions for the impaired molecular weight experimentally. Therefore, posttranslational modifications and some not controlled or biased experimental conditions may lead to the outcome. The ratios of observed molecular weight to expected molecular weight of KNL2 and KNL2Nterm are higher than the ratio of KNL2Cterm (Table 2). This indicates a correlation between the N-terminal part and the increased molecular weight. However, the molecular weight of the KNL2Cterm band differs among the experiments (see 4.1.4.1). To solve this discrepancies we would suggest to ensure the expression of KNL2 protein variants in inclusion bodies by raising the expression temperature and inducer concentration followed by isolating, washing, denaturating and renaturating inclusion body proteins (Simpson, 2003; Green and Sambrook, 2012). Proteins captured in inclusion bodies shall be protected from posttranslational modification (Sahdev et al., 2008). If the size discrepancy would be still detected, changing the expression vector and the host strain may overcome possible experimental bias.

5.2.2 C-terminal truncation of KNL2 variants might be caused by proteolysis or translational error

The designated main protein bands of KNL2 variants consist of estimated 10-70% of the whole protein purification depending on each variant (see Figure 4 & Supplemental Figure 12). Stable KNL2 protein variants were obtained, however, optimization is required. The truncation of the KNL2 protein variants happens at the C-terminal site, because the N-terminal His tag was intact and detected by the anti-His tag antibodies (see 4.1.4.2). This truncation may have two possible sources: the activity of bacterial C-terminus specific endoproteases or exopeptidases (Miller, 1975) or 'codon usage bias' (Green and Sambrook, 2012).

BL21 strain is deficient in the proteases *lon* and *ompT* to decrease the degradation of expressed recombinant proteins (Green and Sambrook, 2012). However, *E. coli* cells contains more than 20 proteolytic enzymes (Murby et al., 1996) and they are needed to maintain a physiological protein level and cellular homeostasis. Therefore, the C-terminal degradation of recombinant proteins by *E. coli* enzymes like HflB, Clp (Herman et al., 1998) or dcp proteases (Henrich et al., 1993) may occur. However, the non-polar character of last five amino acids of the C-terminus determines the activity of C-terminus specific proteases in *E. coli* (Parsell et al., 1990) and just 2 of the last 5 amino acids of the C-terminus of KNL2 protein variants are non-polar (Supplemental Figure 14). Three approaches can be applied to avoid degradation: I) separating the protein from proteases by isolating inclusion bodies, because recombinant proteins are protected from degradation inside inclusion bodies (Schoner et al., 1985; Wheelwright, 1991), II) separating the protein from proteases the by fusing a leader signal peptide to recombinant proteins to induce their secretion out of the host cell cytoplasm (Murby et al., 1996) and III) adding other protease inhibitors like AEBSF, leupeptin, benzamidine or E-64 together with the serine protease inhibitor PMSF to ensure stable inhibition of serine and cysteine proteases. EDTA inhibits metallo-proteases. However, EDTA cannot be applied, because this chelating agent disturb the down-stream IMAC purification procedure (Green and Sambrook, 2012).

"Codon usage bias […] is usually defined as a species-specific deviation from uniform codon usage in the coding regions of genomic sequences."(Fox and Erill, 2010) If many codons, which are rarely used in *E. coli*, are present in the coding sequence of the inserted eukaryotic genes, translational errors like frame-shifts and truncations will arise (Kane, 1995). The sequences of *KNL2* variants contain around 8% of rare codons (Supplemental Table 3). Even one rare codon might be enough to decrease the efficiency of translation

(Kane, 1995). The codon bias can be overcome by using the Rosetta 2 strain instead of the BL-21 strain, additionally the gene sequences can be optimized for codon usage in *E. coli* by several companies (Green and Sambrook, 2012). These approaches might help to decrease truncation and improve expression of KNL2 variants *in vivo*.

The total protein extracts of induced *E. coli* cells show weaker signals for bands of truncated KNL2 protein variants in the developed Western blot film as the elution fractions of the protein purification (Figure 5). In the total protein extracts protein degradation should be inhibited through the applied sample buffer and sample treatment (Green and Sambrook, 2012). However, some investigators suggest treating the cells with incubation at 85°C for 3 min instead of 15 min incubation at 65°C to ensure the denaturation of all proteases (Grabski and Burgess, 2001). Nevertheless, these results indicate proteolytic degradation during the purification procedure and translation errors are also possible. Therefore, we suppose to use Rosetta 2 strain for protein expression and add additional protease inhibitors during protein purification to gain intact KNL2 protein variants. Non-degraded protein solutions are better binding partner in EMSA experiments, because the whole protein can be necessary to form the protein-DNA complex (Simpson, 2003) (see 5.3.2). Partially truncated protein solutions are sufficient for kinase assays or for antibody production, because KNL2 variant serve as substrates and biological activity must not be alter by degradation (Simpson, 2003).

5.3 Characterization of KNL2 as centromeric protein with indicated DNA binding capability and Aurora 3 phosphorylation sites

5.3.1 KNL2 localizes at chromocenters during interphase detected by specific antibodies

The produced anti-KNL2Cterm antibodies show sensitive reaction in ELISA and Western blot (see 4.2.4.1). The KNL2Cterm protein is better recognized than the control protein by anti-KNL2Cterm antibodies in ELISA and 150 pg of KNL2Cterm and 15 ng of KNL2 were detected in Western blot analysis, whereas KNL2Nterm is not. Additionally dimer formation was observed in the Western blot of the related SDS-PAA gel. Such dimerization in denaturating PAA gels are observed by other investigators (see 5.2.1).

These antibodies showed immunostaining signals at chromocenters and nucleoplasm of meristematic interphase nuclei of *Arabidopsis thaliana* (see 4.2.4.2). Chromocenters are distinct, dense bodies formed by peri-/centromeric heterochromatin, which ensure

centromeric function (Karpen and Allshire, 1997; Tiang et al., 2012). In nuclei of *KNL2* SALK mutants, anti-KNL2 immunostaining signals were detected only in nucleoplasm. We suppose that the nucleoplasmic signal is a result of unspecific staining by KNL2-unspecific antibodies present in the sample of purified IgGs and we can not exclude that the KNL2 specific antibodies can bind multiple epitopes and label nucleoplasmic *Arabidopsis* proteins. Nevertheless, the centromeric signals are specific for wild type. Centromeric signals during interphase are also obtained by Lermontova et al. (2013) using *KNL2-EYFP* or *KNL2*Cterm-*EYFP* fusion constructs stably expressed in *Arabidopsis thaliana*. The signal of anti-KNL2Cterm antibodies co-localize with CenH3 at chromocenters (see 4.2.4.3). This indicates a close spatial relationship and possible interaction between those two proteins. Finally a Western blot with the produced antibodies (see 4.2.4.4) shows the related band of native KNL2. Other bands were also detected. This might have the same reason as the nucleoplasmic staining in the immunostaining experiments. In summary, these results show the antibodies are sensitive and specific and they shall be used in future work like co-immunoprecipitation or ChiP experiments (Golemis (Ed.), 2002; Collas, 2010) and KNL2 is incorporated into centromeric regions.

5.3.2 Characterization of DNA-binding capability of KNL2Cterm in a proposed regulatory network

The interaction between KNL2 protein variants and the cen180, the centromeric repeat sequence of *Arabidopsis*, was verified by EMSA experiments (see 4.2.1). KNL2Cterm induce a shift of the cen180 band indicating a protein-DNA interaction, however, KNL2 and KNL2Nterm do not show such a shift.

The open question is: why does KNL2Cterm shows protein-DNA interaction, whereas KNL2 not? Maybe the experimental biases like the urea concentration, degradation or non-native conformation are the reasons. Thus, KNL2 protein variants must be isolated in a soluble form with native conformation without degradation. Protein expression in eukaryotic expression systems (Rosenberg, 2005; Green and Sambrook, 2012) or following experimental changes may help to increase solubility: lowering of the expression temperature and inducer concentration (Schein and Noteborn, 1988; Green and Sambrook, 2012), altering media composition and host strain (Simpson, 2003; Green and Sambrook, 2012), co-expressing molecular chaperones (Amrein et al., 1995; Murby et al., 1996), fusing the protein to protein tags with high solubility like MBP (Baneyx, 1999) and adding ‘stabilizer’ like sorbitol or glycyl betaine (Blackwell and Horgan, 1991). However,

KNL2Cterm shows DNA binding capability even in 0.5 M urea or in precipitated state of the protein sample, whereas KNL2 and KNL2Nterm do not induce a shift of cen180 under the same conditions. Some investigators show protein-DNA interaction in 0.2 M urea (Cohen et al., 1996). Therefore, we hypothesize that KNL2Nterm may inhibit the DNA binding to give another explanation of the different behavior of KNL2 and KNL2Cterm in the EMSA experiments. Such intramolecular inhibitory mechanism is known in many transcription factors of eukaryotic organisms: mouse Isl1 (Gadd et al., 2013), human CDP/Cux (Truscott et al., 2004), human p63 (Serber et al., 2002), Swi4 in *Saccharomyces cerevisiae* (Baetz and Andrews, 1999) and b-Myb in *Xenopus* (Humbert-Lan and Pieler, 1999). In general "[c]onformational changes, including local protein folding, play important roles in protein-DNA interactions" (Petersen et al., 1995).

We would like to propose a model to the hypothesis to explain the difference between centromeric localization of KNL2 during whole interphase and the assumed 'licensing and priming' activity during G_2 phase (see 1.5). KNL2 is probably not active during G_1 phase and may be regulated by an internal mechanism. According to our model KNL2 is incorporated in the inner kinetochore throughout interphase and inactive during G_1 by an intramolecular inhibitory mechanism. During G_2 phase intramolecular transition stops the inhibition of the C-terminal DNA binding capability. KNL2 is 'anchored' to the DNA and chromatin remodeling complexes like H3K9 methyltransferases (Lermontova et al., 2013) are loaded close to their target DNA sites possibly by the N-terminal SANTA domain. This network based on KNL2 lead to the 'priming and licensing' of CenH3 assembly, which is stopped at the beginning of M phase by the removal of KNL2 from kinetochore. An EMSA experiment with the mixture of KNL2Nterm and KNL2Cterm can verify the hypothesis of the intramolecular regulation (Humbert-Lan and Pieler, 1999). Additionally the EMSA experiment with the DNA binding competitor (Figure 13) has to be repeated to verify the specificity of the DNA binding capability.

Then, the question arises: how is the nuclear deposition of KNL2 during interphase and the absence during M phase determined (Lermontova et al., 2013) (see 4.2.4.2 & 5.3.1)? Life cell imaging experiments of *Arabidopsis thaliana* root tip cells stably transformed with the *EYFP-KNL2*/*KNL2*Nterm/*KNL2*Cterm fusion construct showed no fluorescence during mitosis (Lermontova et al., 2013). Therefore, the absence of KNL2 at chromocenters during M phase is independent of the N-/or C-terminus. The DNA binding capability of human KNL2 C-terminus is recently described as non-specific to a sequence motif and interacting with DNA in a novel fashion (Maddox, 2013). These findings support the

'anchoring' function of the C-terminus. If this mechanism is also conserved in *Arabidopsis* KNL2 C-terminus, the regulation of nuclear localization of KNL2 at chromocenters might be not determined by the C-terminal DNA binding alone. In the next chapter (see 5.3.3.3), we suppose such regulatory network for KNL2 localization.

Two unexpected findings of the EMSA experiments shall be discussed. 1) KNL2Nterm dissolved in 0.19 M urea showed a signal reduction of cen180 band at 180 bp indicating some protein-DNA interaction (Supplemental Figure 10). Soluble KNL2Nterm protein solution without urea is needed to verify this effect. 2) In the same experiment KNL2Cterm containing sample without urea shows more stretched His tag signal than the protein-DNA sample with 0.5 M urea. This indicates a better mobility of the protein-DNA complex. The urea concentration might not influence the shift of cen180, however, it affect the stability of protein-DNA complex during electrophoresis.

5.3.3 Phosphorylation of KNL2 by Aurora kinase 3 and a proposed KNL2 localization model

5.3.3.1 Introduction of putative KNL2 kinases

The localization of *Arabidopsis* KNL2 at the kinetochore might depend on phosphorylation (see 1.6.2). Four candidate kinases of KNL2 are proposed by *in silico* co-expression analysis (see 4.2.3.1): Aurora kinases 1, 2 and 3; and Haspin kinase.

5.3.3.1.1 Aurora kinases in *Arabidopsis thaliana*

The three types of *Arabidopsis* Aurora kinases share a conserved enzymatic Ser/Thr kinase domain (Demidov et al., 2005; Kawabe et al., 2005). Aurora kinases phosphorylate histone 3 at serine 10 (Demidov et al., 2005; Petrovska et al., 2012). This is an important epigenetic mark, which often occurs with heterochromatic H3K9me2 modification (see 1.3) at pericentromeric sites in cell cycle dependent manner in mitosis (Gernand et al., 2003; Fuchs et al., 2006; Eberlin et al., 2008). Furthermore, H3S10ph may have other functional importance. For instance, it negatively affects H3K9 methylation. H2S10ph can be established independently of H3K9 methylation and it is involved in regulation of transcription of centromeric sequences (Demidov et al., 2009). Some investigators suggest a phosphorylation of CenH3 by Aurora kinases (Kurihara et al., 2006). *Arabidopsis* plants with deregulated expression of Aurora kinase show defects in cell division. Overall Aurora kinases have mitotic function, are expressed during mitosis and most abundant in dividing

cells. The mitotic expression differs between the three kinases: *Aurora kinase 1* shows very high (≈6000 normalized signals), *Aurora kinase 3* moderate (≈2000 signals) and *Aurora kinase 2* low mitotic expression (≈1000 signals) (Demidov et al., 2005).

The localization of the Aurora kinases differs during the cell cycle, too. Aurora 1 and 2 are localized at cytoplasm and nucleus during interphase and during M phase both kinases are associated with spindle microtubules (Demidov et al., 2005; Kawabe et al., 2005; Van Damme et al., 2011). However, Aurora kinase 3 has different localization pattern: during interphase it is localized in nucleus and cytoplasma (Demidov et al., 2005; Van Damme et al., 2011) or in cytoplasm and around the nucleus (Kawabe et al., 2005), and during mitosis at distinct spots corresponding peri-/centromeric regions (Demidov et al., 2005; Van Damme et al., 2011).

5.3.3.1.2 Haspin kinase in *Arabidopsis thaliana*

Haspin kinase is another Ser/Thr kinase, which phosphorylates histone H3 at threonine 3 (Ashtiyani et al., 2011). H3T3ph can be found along the entire length of chromosome arms and is associated with chromosome condensation (Houben et al., 2007). H3T3ph and H3S10ph/H3S28ph differ in temporal and spatial phosphorylation pattern in vertebrates and plants indicating diverse regulatory function between Haspin and Aurora kinases (Polioudaki et al., 2004; Houben et al., 2007). In summary, H3S10ph/H3S28ph are associated with chromosome cohesion and are located at pericentromeric sites, whereas H3T3ph play a role in chromosome condensation and is observed over the entire chromosome arms. RNAi experiments of human *Haspin* lead to abnormal metaphase chromosome alignment indicating the role of Haspin in chromosome segregation (Dai and Higgins, 2005). In plants, less H3T3ph modification and no defects in mitotic cell cycle are observed in the *Haspin* RNAi down-regulated cells. However, complete depletion causes defects in embryo development (Ashtiyani et al., 2011). Moreover, Haspin kinase is highly expressed in proliferated tissues, and this kinase is enzymatically active throughout the cell cycle unlike Aurora kinases (Dai and Higgins, 2005; Ashtiyani et al., 2011). Haspin kinase is located in cytoplasm during interphase and invades the nucleus after the nuclear envelope breakdown. After the breakdown Haspin localizes with H3T3ph modifications over chromosomes, but not directly with dynamic mitotic structures (Kurihara et al., 2011). The expression of *Haspin kinase* seem to be controversial: some investigators found mitosis specific expression like Aurora kinase 2 (Kurihara et al., 2011), other report a expression over the whole cell cycle (Houben et al., 2007; Ashtiyani et al., 2011).

5.3.3.2 Interpretation of co-expression data of *KNL2 kinase* candidates

Our candidates are important players for the regulation of cell division regarding their role in cohesion or condensation, microtubule formation or development. We performed a co-expression analysis to pick one candidate for down-stream kinase assay (see 4.2.3.1). Co-expression indicates functional interaction of the transcripts. *Aurora kinase 2* shows the lowest correlation coefficient the *in silico* analysis. *Aurora kinases 1* and *3* as well as *Haspin kinase* have comparable correlation coefficients. Therefore, all these kinases are the primary candidates and are considered for future work. We choose Aurora kinase 3 for kinase assay, because this kinase localize clearly at chromocentric spots during M phase in contrast to the other candidates.

5.3.3.3 Aurora kinase 3 phosphorylates KNL2 and might regulate its localization

Kinase assay showed phosphorylation of KNL2 by Aurora kinase 3 at both N-terminal and C-terminal parts (see 4.2.3.2). Additionally minor signals are gain for auto-phosphorylation. Moreover, KNL2 variants contain one conserved phosphorylation motif in the additional C-terminal part derived from the expression vector (Supplemental Figure 14). However, C-terminal truncated products show also phosphorylation (Figure 16 & Supplemental Figure 13, see 5.2.2). Therefore, co-immunoprecipitation of native KNL2 and Aurora kinase 3 is recommended to verify these results. Nevertheless, the preliminary data show phosphorylation of KNL2 by Aurora kinase 3 and we want to propose following model. Aurora kinase 3 localizes at chromocenters after the nuclear envelope breakdown (see 5.3.3.1.1) and phosphorylates KNL2. Due to this phosphorylation KNL2 is removed from chromocenters (Lermontova et al., 2013) (see 1.6.2 & 5.3.2). After M phase Aurora kinase 3 leaves the centromeric positions (Kawabe et al., 2005) and KNL2 bind again inner kinetochore structures. Therefore, we suggest Aurora kinase 3 is involved in the localization of KNL2 during cell cycle.

Future work with application of Aurora kinase specific inhibitors and knock-out or knock-down of *Aurora kinases*' expression as well as the analysis of phosphorylated sites by peptides based kinase assay (Demidov et al., 2009; Lermontova et al., 2013), can help to verify the mechanism and functional importance of KNL2 phosphorylation by Aurora kinase 3.

5.4 Conclusions

CenH3 assembly factor KNL2 was identified in *Arabidopsis thaliana* by our group (Lermontova et al., 2013). Recombinant KNL2 protein variants (full-length, N- and C-terminal parts) were successfully produced in *E. coli* and anti-KNL2 antibodies were successfully generated in rabbits against the C-terminal part of KNL2. Anti-KNL2 antibodies showed immunostaining signals within nucleus and at chromocenters of meristematic interphase nuclei. No signals were detected at centromeres during mitosis. Double immunostaining experiments showed co-localization of CenH3 and KNL2 at chromocenters indicating assembly factor characteristics. C-terminal part of KNL2 showed interaction with 180 bp centromeric repeat sequence. Moreover, the specificity of this DNA binding has to be verified. An unspecific DNA binding capability would resemble the human homologue function. Aurora kinase 3 phosphorylates KNL2 at both N- and C-terminal parts. This phosphorylation might regulate the localization of KNL2 during the cell cycle.

5.5 Outlook

Modification and optimization of our experiments is required to increase the solubility of KNL2 protein variants (see 4.1.2.2), decrease the truncation (see 5.2.2), evaluate the specificity of the DNA binding capability of C-terminal part of KNL2 (see 5.3.2) and check additional kinases for KNL2 phosphorylation (see 5.3.3.2). In short, we would like to express the KNL2 protein variants in a new expression system to tag the recombinant proteins with a solubility mediating MBP. The degradation can be controlled by secreting the recombinant proteins out of the cells or using additional protease inhibitors. Then, the EMSA experiments will be repeated with soluble proteins and a South-Western blot can be applied for verification. A suggested intramolecular inhibition mechanism can be controlled by mixing KNL2Nterm and KNL2Cterm (see 5.3.2). Remaining anti-KNL2 antisera shall be purified by an antigen affinity chromatography and the Western blot experiments (see 4.2.4.4) can be modified to an antigen competition assay to verify the wild type specific KNL2 band. Several approaches can be applied to get further insights of the interaction between KNL2 and KNL2 kinases: a kinase assay with peptides of the predicted target sites, RNAi approaches of *KNL2 kinases* in *Arabidopsis thaliana*, kinase inhibitor treatment of *Arabidopsis* seedlings and a BiFC assay including KNL2 and the related kinases. Also experimental biases appearing at certain experiments have to be

verified (see 4.1.2.2, 4.1.4.1, 4.2.1, 4.2.2, 5.2.1 & 5.3.2). Additional experiments such as ChiP or co-immunoprecipitation are also required to get further insight in the biological function of KNL2 in *Arabidopsis* (see 4.2.2, 5.3.1 & 5.3.3.3).

Finally we want to suggest rather long term projects, which have practical implications. The research on centromere and kinetochore proteins like KNL2 can provide the knowledge to construct minichromosomes as artificial chromosome cloning vectors, which can be used to manipulate crops (Sullivan et al., 2001). Furthermore, CenH3 is connected to the production of double haploid plants, because crossings between wild type and *CenH3* null mutant can generate haploid progenies (Ravi and Chan, 2010). Double haploid plants play an important role in plant breeding. The possible connection between CenH3 and KNL2 provide the assumption, that *KNL2* as an upstream component of the CenH3 assembly mechanism can substitute *CenH3* in this context.

6 Acknowledgments

I would like to express my gratitude to my supervisor Dr. Inna Lermontova. She gave me the opportunity to work at IPK research group and supports me the whole time of my work. It is a pleasure to work together with her. And I appreciate our stimulating discussions and I also want to thank her especially for supporting me in writing and correcting this thesis.

All collaborators helped me a lot and I want to thank all of them. The collaborators are: Dr. Dmitri Demidov, Eva Tomaŝtiková, Dr. Andrea Matros, Dr. Veit Schubert, Ulrike Gresch, Dr. Udo Conrad and Isolde Tillack. Especially I am particularly grateful to Dr. Dmitri Demidov for all his support.

Furthermore, I am thankful to the technicians Joachim Bruder and Andrea Kunze. Joachim helped me with all kinds of media and Andrea gave me support at several techniques.

I want to thank all 'KTE' group members: Dr. Stefan Heckmann, Henriette Leps, Tran Duc Trung, Inge Sporleder, Martina Kühne, Christa Fricke, Dr. Jörg Fuchs, Dr. Xuan Hieu Cao and Dr. Thi Ha Giang Vu. I enjoy the friendly and open atmosphere in the research group.

Moreover, I would like to thank all members of the 'Chromosome Structure and Function' and 'Quantitative Genetics' research group. I am especially indebted to Dr. Andreas Houben and Prof. Jochen Reif. Both supported me in different situations.

I also want to thank my best friends, Mandy Spörl and Christian Müller, without their support not only working and studying is much more difficult.

Without the help of my family I would not come so far. I am grateful to my parents, my brothers and all relatives I know.

I want to thank to whole IPK. This is a really nice place to work.

Finally, I want to give my special thanks to Prof. Ingo Schubert. He was the first person I contacted and helped me to start the work at IPK.

7 Reference

Ahmad, K. and S. Henikoff (2001). "Centromeres are specialized replication domains in heterochromatin." J Cell Biol **153**(1): 101-110.

Alberts, B., A. Johnson, et al. (2002). Molecular Biology of the Cell. New York, Garland Science.

Amrein, K. E., B. Takacs, et al. (1995). "Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an Escherichia coli expression system overproducing the bacterial chaperones GroES and GroEL." Proc Natl Acad Sci U S A **92**(4): 1048-1052.

Annunziato, A. T. (2005). "Split decision: what happens to nucleosomes during DNA replication?" J Biol Chem **280**(13): 12065-12068.

Aon, J. C., R. J. Caimi, et al. (2008). "Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of Escherichia coli." Appl Environ Microbiol **74**(4): 950-958.

Ashtiyani, R. K., A. M. B. Moghaddam, et al. (2011). "AtHaspin phosphorylates histone H3 at threonine 3 during mitosis and contributes to embryonic patterning in Arabidopsis." Plant Journal **68**(3): 443-454.

Baetz, K. and B. Andrews (1999). "Regulation of cell cycle transcription factor Swi4 through auto-inhibition of DNA binding." Mol Cell Biol **19**(10): 6729-6741.

Baneyx, F. (1999). "Recombinant protein expression in Escherichia coli." Curr Opin Biotechnol **10**(5): 411-421.

Barnhart, M. C., P. H. Kuich, et al. (2011). "HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore." J Cell Biol **194**(2): 229-243.

Bernatavichute, Y. V., X. Zhang, et al. (2008). "Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana." PLoS One **3**(9): e3156.

Bjerrum, O. J. and C. Schafer-Nielsen (1986). Analytical Electrophoresis. Weinheim, Verlag Chemie.

Black, B. E., L. E. Jansen, et al. (2007). "Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain." Mol Cell **25**(2): 309-322.

Blackwell, J. R. and R. Horgan (1991). "A novel strategy for production of a highly expressed recombinant protein in an active form." FEBS Lett **295**(1-3): 10-12.

Blower, M. D. and G. H. Karpen (2001). "The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions." Nat Cell Biol **3**(8): 730-739.

Boyer, L. A., R. R. Latek, et al. (2004). "The SANT domain: a unique histone-tail-binding module?" Nat Rev Mol Cell Biol **5**(2): 158-163.

Brown, W. and C. Tyler-Smith (1995). "Centromere activation." Trends Genet **11**(9): 337-339.

Bui, M., E. K. Dimitriadis, et al. (2012). "Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo." Cell **150**(2): 317-326.

Carroll, C. W., K. J. Milks, et al. (2010). "Dual recognition of CENP-A nucleosomes is required for centromere assembly." J Cell Biol **189**(7): 1143-1155.

Chen, Y., R. E. Baker, et al. (2000). "The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain." Mol Cell Biol **20**(18): 7037-7048.

Cohen, D. M., S. R. Gullans, et al. (1996). "Urea inducibility of egr-1 in murine inner medullary collecting duct cells is mediated by the serum response element and adjacent Ets motifs." Journal of Biological Chemistry **271**(22): 12903-12908.
Collas, P. (2010). "The current state of chromatin immunoprecipitation." Mol Biotechnol **45**(1): 87-100.
Cooper, G. M. (2000). The Cell: A Molecular Approach, Sinauer Associates.
Cooper, J. L. and S. Henikoff (2004). "Adaptive evolution of the histone fold domain in centromeric histones." Mol Biol Evol **21**(9): 1712-1718.
Copenhaver, G. P., K. Nickel, et al. (1999). "Genetic definition and sequence analysis of Arabidopsis centromeres." Science **286**(5449): 2468-2474.
Dai, J. and J. M. Higgins (2005). "Haspin: a mitotic histone kinase required for metaphase chromosome alignment." Cell Cycle **4**(5): 665-668.
Dalal, Y., T. Furuyama, et al. (2007). "Structure, dynamics, and evolution of centromeric nucleosomes." Proc Natl Acad Sci U S A **104**(41): 15974-15981.
Dalal, Y., H. Wang, et al. (2007). "Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells." PLoS Biol **5**(8): e218.
De Rop, V., A. Padeganeh, et al. (2012). "CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly." Chromosoma **121**(6): 527-538.
Dell, A., A. Galadari, et al. (2010). "Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes." Int J Microbiol **2010**: 148178.
Demidov, D., S. Hesse, et al. (2009). "Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in Arabidopsis." Plant J **59**(2): 221-230.
Demidov, D., D. Van Damme, et al. (2005). "Identification and dynamics of two classes of aurora-like kinases in Arabidopsis and other plants." Plant Cell **17**(3): 836-848.
Dephoure, N., C. Zhou, et al. (2008). "A quantitative atlas of mitotic phosphorylation." Proc Natl Acad Sci U S A **105**(31): 10762-10767.
Depinet, T. W., J. L. Zackowski, et al. (1997). "Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA." Hum Mol Genet **6**(8): 1195-1204.
Dunleavy, E. M., G. Almouzni, et al. (2011). "H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase." Nucleus **2**(2): 146-157.
Eberlin, A., C. Grauffel, et al. (2008). "Histone H3 tails containing dimethylated lysine and adjacent phosphorylated serine modifications adopt a specific conformation during mitosis and meiosis." Mol Cell Biol **28**(5): 1739-1754.
Fang, Y. and D. L. Spector (2005). "Centromere positioning and dynamics in living Arabidopsis plants." Mol Biol Cell **16**(12): 5710-5718.
Fox, J. M. and I. Erill (2010). "Relative codon adaptation: a generic codon bias index for prediction of gene expression." DNA Res **17**(3): 185-196.
Fuchs, J., D. Demidov, et al. (2006). "Chromosomal histone modification patterns--from conservation to diversity." Trends Plant Sci **11**(4): 199-208.
Fujita, Y., T. Hayashi, et al. (2007). "Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1." Dev Cell **12**(1): 17-30.
Gadd, M. S., D. A. Jacques, et al. (2013). "A structural basis for the regulation of the LIM-homeodomain protein Islet 1 (Isl1) by intra- and intermolecular interactions." J Biol Chem.
Gernand, D., D. Demidov, et al. (2003). "The temporal and spatial pattern of histone H3 phosphorylation at serine 28 and serine 10 is similar in plants but differs between mono- and polycentric chromosomes." Cytogenet Genome Res **101**(2): 172-176.

Gherardi, E. and A. Coffer (1991). "Purification and characterization of scatter factor." EXS **59**: 53-62.
Golemis (Ed.), E. (2002). Protein-Protein Interactions:A molecular cloning manual. New York, Cold Spring Harbor Laboratory Press.
Grabski, A. C. and R. R. Burgess (2001). "Preparation of protein samples for SDS-polyacrylamide gel electrophoresis: procedures and tips." inNovations **13**: 10-12.
Granger-Schnarr, M., R. Lloubes, et al. (1988). "Specific protein-DNA complexes: immunodetection of the protein component after gel electrophoresis and Western blotting." Anal Biochem **174**(1): 235-238.
Green, M. R. and J. Sambrook (2012). Molecular cloning. New York, Cold Spring Harbor Laboratory Press.
Hauptmann, V., N. Weichert, et al. (2013). "Native-sized spider silk proteins synthesized in planta via intein-based multimerization." Transgenic Res **22**(2): 369-377.
Hellman, L. M. and M. G. Fried (2007). "Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions." Nat Protoc **2**(8): 1849-1861.
Henikoff, S., K. Ahmad, et al. (2001). "The centromere paradox: stable inheritance with rapidly evolving DNA." Science **293**(5532): 1098-1102.
Henrich, B., S. Becker, et al. (1993). "dcp gene of Escherichia coli: cloning, sequencing, transcript mapping, and characterization of the gene product." J Bacteriol **175**(22): 7290-7300.
Herman, C., D. Thevenet, et al. (1998). "Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH)." Genes Dev **12**(9): 1348-1355.
Houben, A., D. Demidov, et al. (2007). "Phosphorylation of histone H3 in plants--a dynamic affair." Biochim Biophys Acta **1769**(5-6): 308-315.
Howman, E. V., K. J. Fowler, et al. (2000). "Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice." Proc Natl Acad Sci U S A **97**(3): 1148-1153.
Humbert-Lan, G. and T. Pieler (1999). "Regulation of DNA binding activity and nuclear transport of B-Myb in Xenopus oocytes." J Biol Chem **274**(15): 10293-10300.
Isbert, S., K. Wagner, et al. (2012). "APP dimer formation is initiated in the endoplasmic reticulum and differs between APP isoforms." Cell Mol Life Sci **69**(8): 1353-1375.
Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli." Curr Opin Biotechnol **6**(5): 494-500.
Karpen, G. H. and R. C. Allshire (1997). "The case for epigenetic effects on centromere identity and function." Trends Genet **13**(12): 489-496.
Kawabe, A., S. Matsunaga, et al. (2005). "Characterization of plant Aurora kinases during mitosis." Plant Mol Biol **58**(1): 1-13.
Kim, I. S., M. Lee, et al. (2012). "Roles of Mis18alpha in epigenetic regulation of centromeric chromatin and CENP-A loading." Mol Cell **46**(3): 260-273.
Kurihara, D., S. Matsunaga, et al. (2006). "Aurora kinase is required for chromosome segregation in tobacco BY-2 cells." Plant J **48**(4): 572-580.
Kurihara, D., S. Matsunaga, et al. (2011). "Identification and characterization of plant Haspin kinase as a histone H3 threonine kinase." Bmc Plant Biology **11**.
Kyhse-Andersen, J. (1984). "Electroblotting of Multiple Gels - a Simple Apparatus without Buffer Tank for Rapid Transfer of Proteins from Polyacrylamide to Nitrocellulose." Journal of Biochemical and Biophysical Methods **10**(3-4): 203-209.
Lagana, A., J. F. Dorn, et al. (2010). "A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A." Nat Cell Biol **12**(12): 1186-1193.

Lermontova, I., O. Koroleva, et al. (2011). "Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation." Plant J **68**(1): 40-50.
Lermontova, I., M. Kuhlmann, et al. (2013). "Arabidopsis KNL2 is an upstream component for cenH3 deposition at centromeres." Plant Cell (accepted).
Lermontova, I., T. Rutten, et al. (2011). "Deposition, turnover, and release of CENH3 at Arabidopsis centromeres." Chromosoma **120**(6): 633-640.
Lermontova, I. and I. Schubert (2013). Plant Centromere Biology, John Wiley & Sons**:** 67-82.
Lermontova, I., V. Schubert, et al. (2006). "Loading of Arabidopsis centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain." Plant Cell **18**(10): 2443-2451.
Lippman, Z., A. V. Gendrel, et al. (2004). "Role of transposable elements in heterochromatin and epigenetic control." Nature **430**(6998): 471-476.
Lipsick, J. S. (1996). "One billion years of Myb." Oncogene **13**(2): 223-235.
Lo, A. W., D. J. Magliano, et al. (2001). "A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA." Genome Res **11**(3): 448-457.
Lodish, H., A. Berk, et al. (2000). Molecular Cell Biology. New York, W. H. Freeman.
Maddox, P. S., F. Hyndman, et al. (2007). "Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin." J Cell Biol **176**(6): 757-763.
Maddox, S. (2013). Centromere epigenome stability is mediated by KNL-2 structural recognition of CENP-A chromatin. Dynamic Kinetochore Workshop. Porto, IBMC**:** 25.
Maggert, K. A. and G. H. Karpen (2000). "Acquisition and metastability of centromere identity and function: sequence analysis of a human neocentromere." Genome Res **10**(6): 725-728.
Meluh, P. B., P. Yang, et al. (1998). "Cse4p is a component of the core centromere of Saccharomyces cerevisiae." Cell **94**(5): 607-613.
Mendiburo, M. J., J. Padeken, et al. (2011). "Drosophila CENH3 is sufficient for centromere formation." Science **334**(6056): 686-690.
Miller, C. G. (1975). "Peptidases and proteases of Escherichia coli and Salmonella typhimurium." Annu Rev Microbiol **29**: 485-504.
Moraes, I. C., I. Lermontova, et al. (2011). "Recognition of A. thaliana centromeres by heterologous CENH3 requires high similarity to the endogenous protein." Plant Mol Biol **75**(3): 253-261.
Moree, B., C. B. Meyer, et al. (2011). "CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly." J Cell Biol **194**(6): 855-871.
Murby, M., M. Uhlen, et al. (1996). "Upstream strategies to minimize proteolytic degradation upon recombinant production in Escherichia coli." Protein Expr Purif **7**(2): 129-136.
Nagaki, K., P. B. Talbert, et al. (2003). "Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres." Genetics **163**(3): 1221-1225.
Naryzhnyi, S. N. (1992). "[Detection of DNA-binding proteins in polyacrylamide gel after denaturing electrophoresis]." Mol Biol (Mosk) **26**(2): 307-314.
Nasuda, S., S. Hudakova, et al. (2005). "Stable barley chromosomes without centromeric repeats." Proc Natl Acad Sci U S A **102**(28): 9842-9847.

Nechemia-Arbely, Y., D. Fachinetti, et al. (2012). "Replicating centromeric chromatin: spatial and temporal control of CENP-A assembly." Exp Cell Res **318**(12): 1353-1360.
Nousiainen, M., H. H. Sillje, et al. (2006). "Phosphoproteome analysis of the human mitotic spindle." Proc Natl Acad Sci U S A **103**(14): 5391-5396.
Oegema, K., A. Desai, et al. (2001). "Functional analysis of kinetochore assembly in Caenorhabditis elegans." J Cell Biol **153**(6): 1209-1226.
Palecek, E., M. Brazdova, et al. (1999). "Effect of transition metals on binding of p53 protein to supercoiled DNA and to consensus sequence in DNA fragments." Oncogene **18**(24): 3617-3625.
Panchenko, T., T. C. Sorensen, et al. (2011). "Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini." Proc Natl Acad Sci U S A **108**(40): 16588-16593.
Parsell, D. A., K. R. Silber, et al. (1990). "Carboxy-terminal determinants of intracellular protein degradation." Genes Dev **4**(2): 277-286.
Peacock, W. J., E. S. Dennis, et al. (1981). "Highly repeated DNA sequence limited to knob heterochromatin in maize." Proc Natl Acad Sci U S A **78**(7): 4490-4494.
Pereira, S. F., L. Goss, et al. (2011). "Eukaryote-like serine/threonine kinases and phosphatases in bacteria." Microbiol Mol Biol Rev **75**(1): 192-212.
Perpelescu, M. and T. Fukagawa (2011). "The ABCs of CENPs." Chromosoma **120**(5): 425-446.
Petersen, J. M., J. J. Skalicky, et al. (1995). "Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix." Science **269**(5232): 1866-1869.
Petrovska, B., V. Cenklova, et al. (2012). "Plant Aurora kinases play a role in maintenance of primary meristems and control of endoreduplication." New Phytol **193**(3): 590-604.
Pluta, A. F., A. M. Mackay, et al. (1995). "The centromere: hub of chromosomal activities." Science **270**(5242): 1591-1594.
Polioudaki, H., Y. Markaki, et al. (2004). "Mitotic phosphorylation of histone H3 at threonine 3." FEBS Lett **560**(1-3): 39-44.
Ravi, M. and S. W. Chan (2010). "Haploid plants produced by centromere-mediated genome elimination." Nature **464**(7288): 615-618.
Ravi, M., F. Shibata, et al. (2011). "Meiosis-specific loading of the centromere-specific histone CENH3 in Arabidopsis thaliana." PLoS Genet **7**(6): e1002121.
Rosenberg, I. M. (2005). Protein Analysis and Purification: Benchtop Techniques. Boston, Birkhäuser.
Roudier, F., I. Ahmed, et al. (2011). "Integrative epigenomic mapping defines four main chromatin states in Arabidopsis." EMBO J **30**(10): 1928-1938.
Sahdev, S., S. K. Khattar, et al. (2008). "Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies." Mol Cell Biochem **307**(1-2): 249-264.
Sardon, T., R. A. Pache, et al. (2010). "Uncovering new substrates for Aurora A kinase." EMBO Rep **11**(12): 977-984.
Schein, C. H. and M. H. M. Noteborn (1988). "Formation of Soluble Recombinant Proteins in Escherichia-Coli Is Favored by Lower Growth Temperature." Bio-Technology **6**(3): 291-294.
Schoner, R. G., L. F. Ellis, et al. (1985). "Isolation and Purification of Protein Granules from Escherichia-Coli-Cells Overproducing Bovine Growth-Hormone." Bio-Technology **3**(2): 151-154.

Serber, Z., H. C. Lai, et al. (2002). "A C-terminal inhibitory domain controls the activity of p63 by an intramolecular mechanism." Mol Cell Biol **22**(24): 8601-8611.
Shelby, R. D., K. Monier, et al. (2000). "Chromatin assembly at kinetochores is uncoupled from DNA replication." J Cell Biol **151**(5): 1113-1118.
Shelby, R. D., O. Vafa, et al. (1997). "Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites." J Cell Biol **136**(3): 501-513.
Shi, J. and R. K. Dawe (2006). "Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27." Genetics **173**(3): 1571-1583.
Shibata, F. and M. Murata (2004). "Differential localization of the centromere-specific proteins in the major centromeric satellite of Arabidopsis thaliana." J Cell Sci **117**(Pt 14): 2963-2970.
Sidorova, N. Y., S. Hung, et al. (2010). "Stabilizing labile DNA-protein complexes in polyacrylamide gels." Electrophoresis **31**(4): 648-653.
Simpson, R. J. (2003). Proteins and proteomics: a laboratory manual. New York, Cold Spring Harbor Laboratory Press.
Steiner, N. C. and L. Clarke (1994). "A Novel Epigenetic Effect Can Alter Centromere Function in Fission Yeast." Cell **79**(5): 865-874.
Stryer, L. (1995). Biochemistry. New York, W.H. Freeman.
Sullivan, B. A., M. D. Blower, et al. (2001). "Determining centromere identity: cyclical stories and forking paths." Nat Rev Genet **2**(8): 584-596.
Sullivan, B. A., S. Schwartz, et al. (1996). "Centromeres of human chromosomes." Environ Mol Mutagen **28**(3): 182-191.
Sullivan, K. F. (2001). "A solid foundation: functional specialization of centromeric chromatin." Curr Opin Genet Dev **11**(2): 182-188.
Tachiwana, H., W. Kagawa, et al. (2011). "Crystal structure of the human centromeric nucleosome containing CENP-A." Nature **476**(7359): 232-235.
Teo, C. H., I. Lermontova, et al. (2013). "De novo generation of plant centromeres at tandem repeats." Chromosoma **122**(3): 233-241.
Tiang, C. L., Y. He, et al. (2012). "Chromosome organization and dynamics during interphase, mitosis, and meiosis in plants." Plant Physiol **158**(1): 26-34.
Truscott, M., L. Raynal, et al. (2004). "The N-terminal region of the CCAAT displacement protein (CDP)/Cux transcription factor functions as an autoinhibitory domain that modulates DNA binding." Journal of Biological Chemistry **279**(48): 49787-49794.
Van Damme, D., B. De Rybel, et al. (2011). "Arabidopsis alpha Aurora kinases function in formative cell division plane orientation." Plant Cell **23**(11): 4013-4024.
van der Ploeg, M., P. van Duijn, et al. (1974). "High-resolution scanning-densitometry of photographic negatives of human metaphase chromosomes. II. Feulgen-DNA measurements." Histochemistry **42**(1): 31-46.
Veiseth, S. V., M. A. Rahman, et al. (2011). "The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis." PLoS Genet **7**(3): e1001325.
Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-1837.
Walter, M., C. Chaban, et al. (2004). "Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation." Plant J **40**(3): 428-438.
Warburton, P. E., C. A. Cooke, et al. (1997). "Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres." Curr Biol **7**(11): 901-904.
Weber, K., J. R. Pringle, et al. (1972). "Measurement of molecular weights by electrophoresis on SDS-acrylamide gels." Methods in Enzymology **26**(1): 2-27.

Wheelwright, S. M. (1991). Protein purification: Design and scale up of Downstream processing. New York, Oxford University Press.
Wong, N. C., L. H. Wong, et al. (2006). "Permissive transcriptional activity at the centromere through pockets of DNA hypomethylation." PLoS Genet **2**(2): e17.
Zhang, D., C. J. Martyniuk, et al. (2006). "SANTA domain: a novel conserved protein module in Eukaryota with potential involvement in chromatin regulation." Bioinformatics **22**(20): 2459-2462.
Zhang, W., H. R. Lee, et al. (2008). "Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize." Plant Cell **20**(1): 25-34.

8 Appendix

KNL2 cDNA:

(KNL2Nterm cDNA:)

ATGACGGAACCAAATCTCGACGAAGATGGTTCCAAGTCGTCTTTCCAGAAAACGGTGGTGCTGAGAGATTGGTGGTTGATAAA
ATGCCCAAAGGAATTCGAAGGGAAACAATTTGGTGTTGCTGGATTCGAAGAGTCCGTCGAGACAAGAGCAATGCGTGTGTTTA
CATCTTCACCAATCACCAAAGCTTTGGATGTTTTCACACTCTTAGCATCTGATGGAATCTATATCACTCTCAGAGGTTTTCTT
AACAAAGAACGCGTTCTTAAAAATGGATTTAACCCTGAGATTTCTCGTGAATTCATCTTCGGGTTTCCTCCTTGTTGGGAACG
AGTTTGTAACAGTTGCTTTGAAGGAGACTCTTTTGGCACTGATGTTAATACCGTCCCTTCGACTATCGAGAAAGCTTGTCCTC
CCATTTTATCACCTTGTAAGTACTCTAACAGGAATCTTAAGGATAATCCAGCTGAGAGCAGAGAGAAAAGCAATGTGACTGAG
ACTGATATTGCAGAGATTAATGATAAAGGTGGTTCTGGAGCAAGAGATATAAAAACTGCAAGGAGAAGGTCTCTTCATCTGCA
AATAAAAAGGATACTTGAATCGAGTAAAGTCCGGAAGACTGCTAATGATGGAGATCATGGTAGTGAATTTTTGAATACGGCTA
AGCGCGGTGATGTAGAAAGAGATGGATGTGAGGTTATCAATAATGAAGACAGTGAATGGAAACTCGATGAAAGTGAAGTCCAG
AATCTTTGTAATGATGGAGATAATGGTAGTGAAGGTTTCATTAAGGCTAAGAGCAGTGATGTAGAAAAAGATAAGAGCGAAGC
TATCGATAATGATGTGATATCTCCAGCGGTTGGAAGTGGTATTAAGCATACCGGTGCAGATAATGTTGATAAAGTGACAAGTG
CAAGTGCTACTGGAGAATCACTTACTTCGGAACAGCAAAATGGTTTACTTGTAACAACAGCATCTCCACATTCCCTGCTTAAA
GACTTAGCCAAGAGTAGCAAACCTGAAAAGAAAGGAATATCCAAGAAAAGTGGCAAGATCCTCAGAAGTGACGACAATGTAGT
AGATCCC

(KNL2Cterm cDNA:)

ATGAATTACTCTGGGACGAAAGTCAAAAGTGCGGAAAACAAAAGGAAAATCGATGCGAGTAAACTCCAGAGTCCGACTAGTAA
TGTTGCAGAACATAGTAAGGAAGGTTTAAATAATGCTAAGAGCAATGACGTAGAAAAAGATGTATGTGTGGCTATCAATAATG
AAGTGATATCACCAGTGAAGGGATTTGGTAAAAGGCTTTCTGGTACAGATGTTGAAAGATTAACAAGTAAGAATGCTACTAAA
GAATCACTGACGTCAGTACAGCGAAAAGGTAGAGTGAAGGTATCGAAGGCATTTCAAGATCCCCTGTCGAAAGGAAAATCAAA
GAAAAGTGAGAAGACCCTTCAAAGTAACAGCAATGTTGTAGAGCCTATGAATCATTTCAGGTCTGAAGCTGAAGAAGCTGAAG
AAAACTTGTCATGGGAAAAAATAAAGAGGAAAATCGACTTTGATGTGGAGGTAACACCGGAAAAAAAAGTGAAGCAGCAGAAG
ACCAATGCGGCGTCTACTGATTCATTGGGACAGAAACGGTCAAGATCAGGAAGGGTGCTTGTGTCATCACTAGAGTTTTGGCG
TAACCAAATTCCTGTTTATGATATGGATCGGAACCTTATCCAAGTAAAGGATGGTAGTGAGACTAACTCCGCTCCATCTAAAG
GAAAAGGATCGGATTCTCGAAAGCGAAGAAACTTGAAAATCAAA

Additional N-terminal sequences with **start codon** and 6x His tag:

ATGTCGTACTAC<u>CATCACCATCACCATCAC</u>CTCGAATCAACAAGTTTGTACAAAAAAGCAGGCTTC

Additional C-terminal sequences with the **stop codon**:

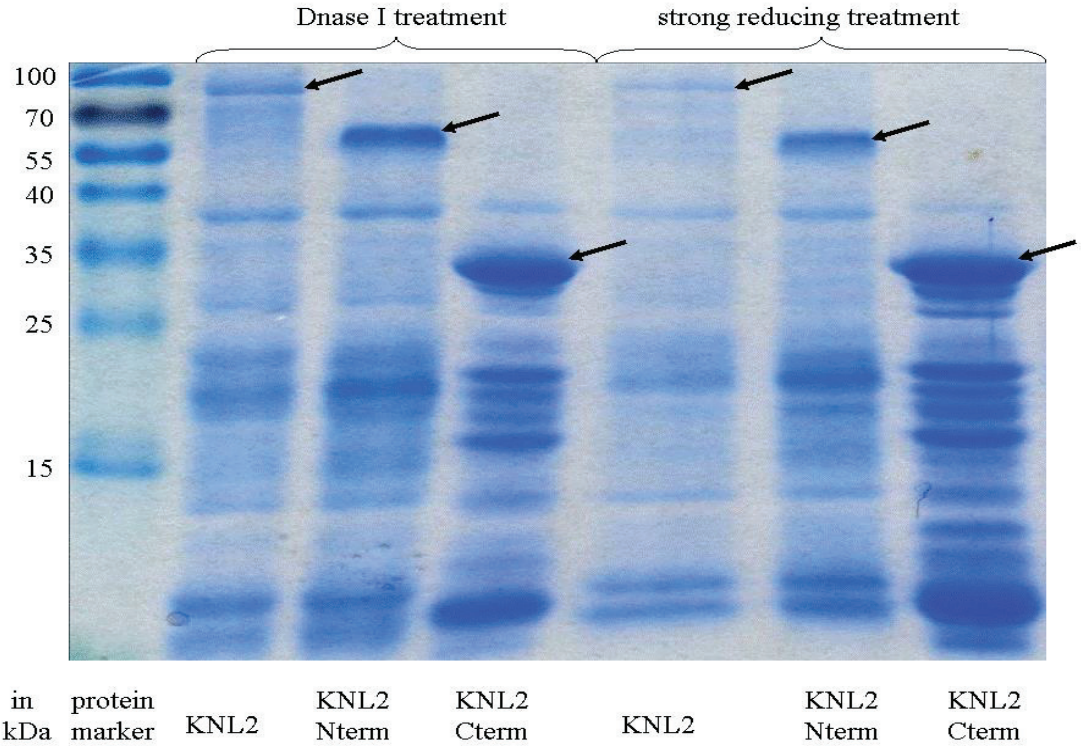
GACCCAGCTTTCTTGTACAAAGTGGTGATTCGAGGCTGC**TAA**

Supplemental Figure 1: The cDNA sequences of *Arabidopsis thaliana KNL2* is annotated in the TAIR database (identifier: AT5G02520) or Uniprot database (identifier: F4KCE9). This sequence served as template for creation of an artificial N-terminal (*KNL2*Nterm) and C-terminal (*KNL2*Cterm) variant. At the site of an internal, in-frame ATG triplet (marked by a green background) *KNL2*Nterm ends and *KNL2*Cterm starts. Due to the insertion into the expression vector pDEST17 additional N- and C-terminal sequences of the vector are part of the recombinant *KNL2* variants (*KNL2*, *KNL2*Nterm, *KNL2*Cterm) along with the corresponding cDNA sequences. The 6x His tag is important for later protein purification (see 2.16).

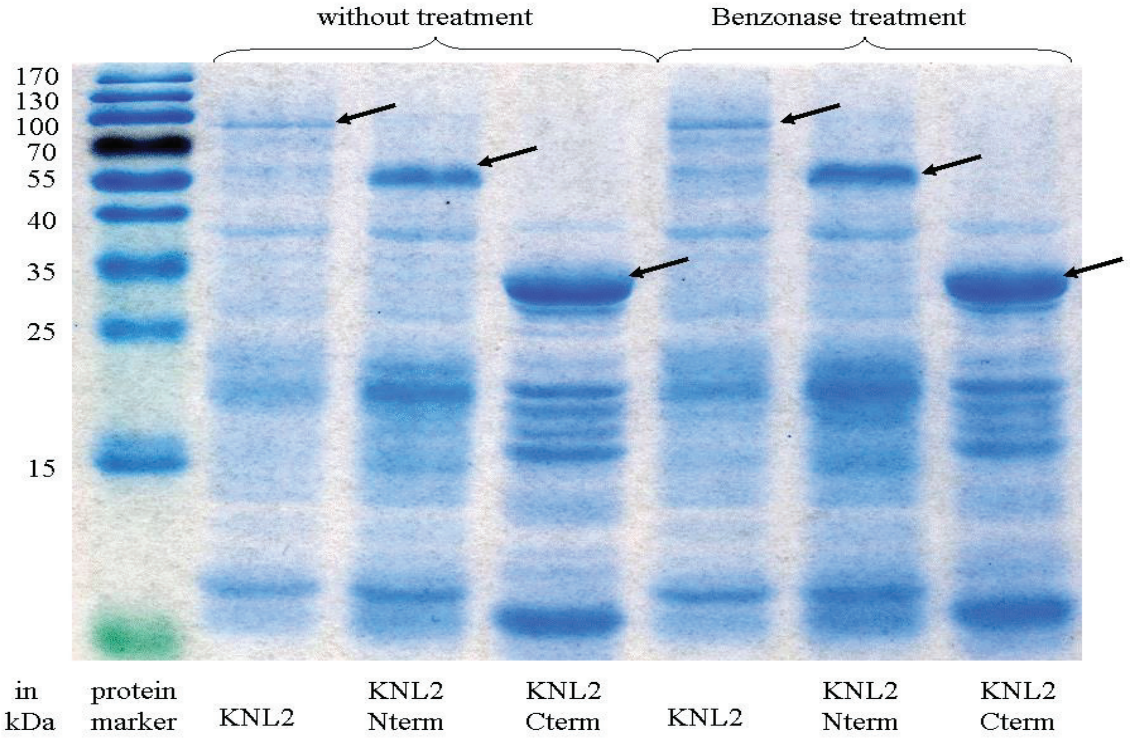
Supplemental Figure 2: Alignment (see 2.10) of the annotated *KNL2* gene (*KNL2*cDNA) with the forward (*KNL2*-fw) and reverse (*KNL2*-rw) sequencing data of isolated expression vector, which carrying the *KNL2* insert. Stars indicate coverage of the three sequences.

Supplemental Figure 3: Alignment (see 2.10) of the N-terminal part of the annotated *KNL2* gene (*KNL2*N) with the forward (*KNL2*N-fw) and reverse (*KNL2*N-rw) sequencing data of isolated expression vector, which carrying the *KNL2*Nterm insert. Stars indicate coverage of the three sequences.

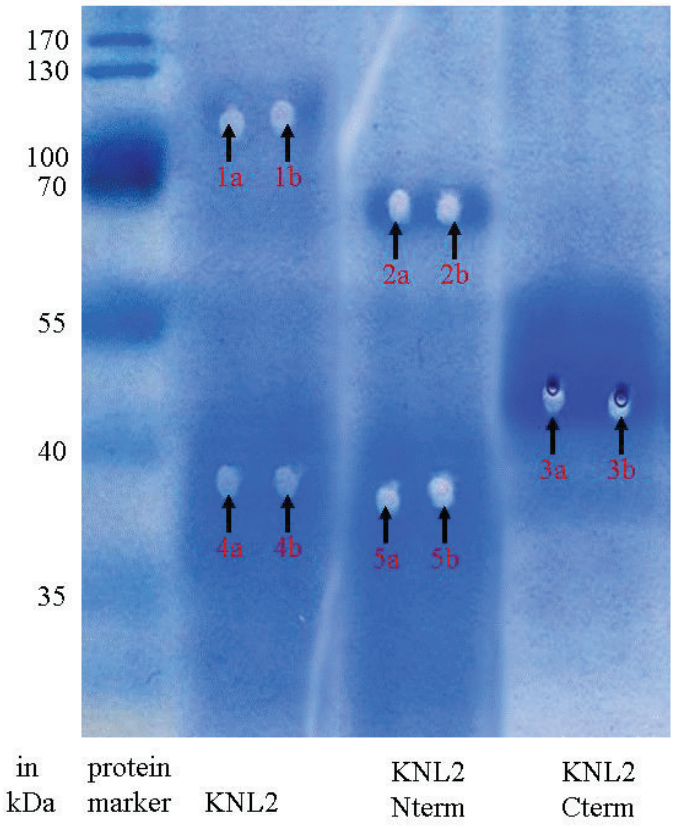
Supplemental Figure 4: Alignment (see 2.10) of the C-terminal part of the annotated *KNL2* gene (*KNL2*C) with the forward (*KNL2*C-fw) and reverse (*KNL2*N-rw) sequencing data of isolated expression vector, which carrying the *KNL2*Cterm insert. Stars indicate coverage of the three sequences.



Supplemental Figure 5: Purified protein solutions were treated with DNase I (see 2.19.1) or strong reducing conditions (see 2.18) and subsequently analyzed in a 12% SDS-PAA gel. The designated bands (arrowed) of the KNL2 variants did not alter in molecular weight after treatment (Figure 4).



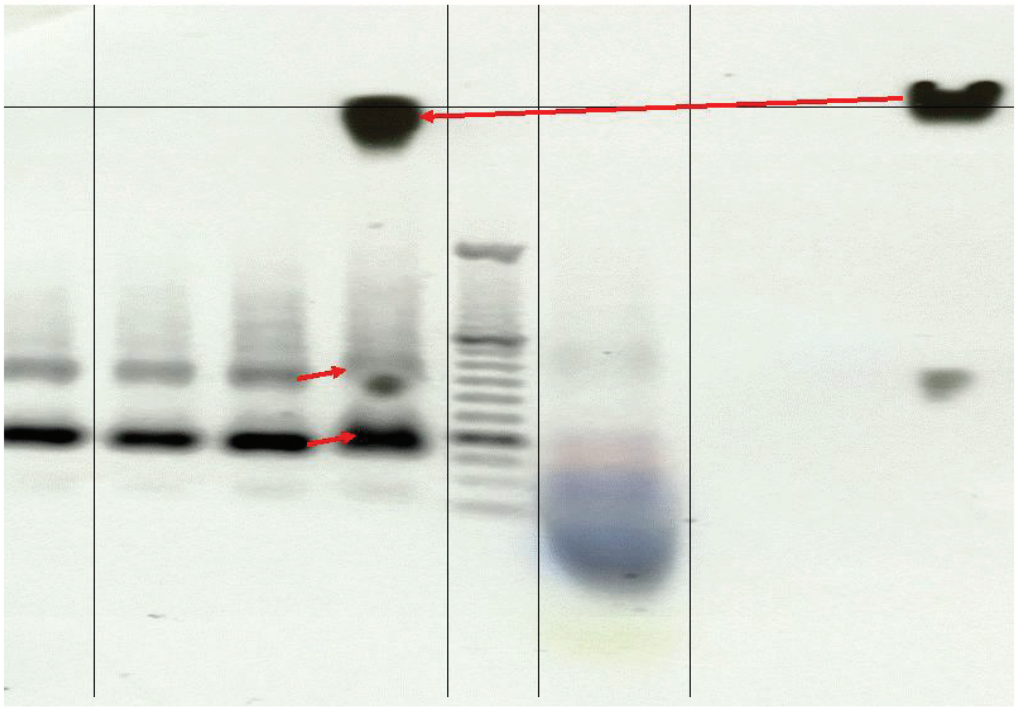
Supplemental Figure 6: Purified protein solutions were treated with Benzonase (see 2.19.2) and subsequently analyzed in a 12% SDS-PAA gel. The designated bands (arrowed) of the KNL2 variants did not alter in the molecular weight after treatment.



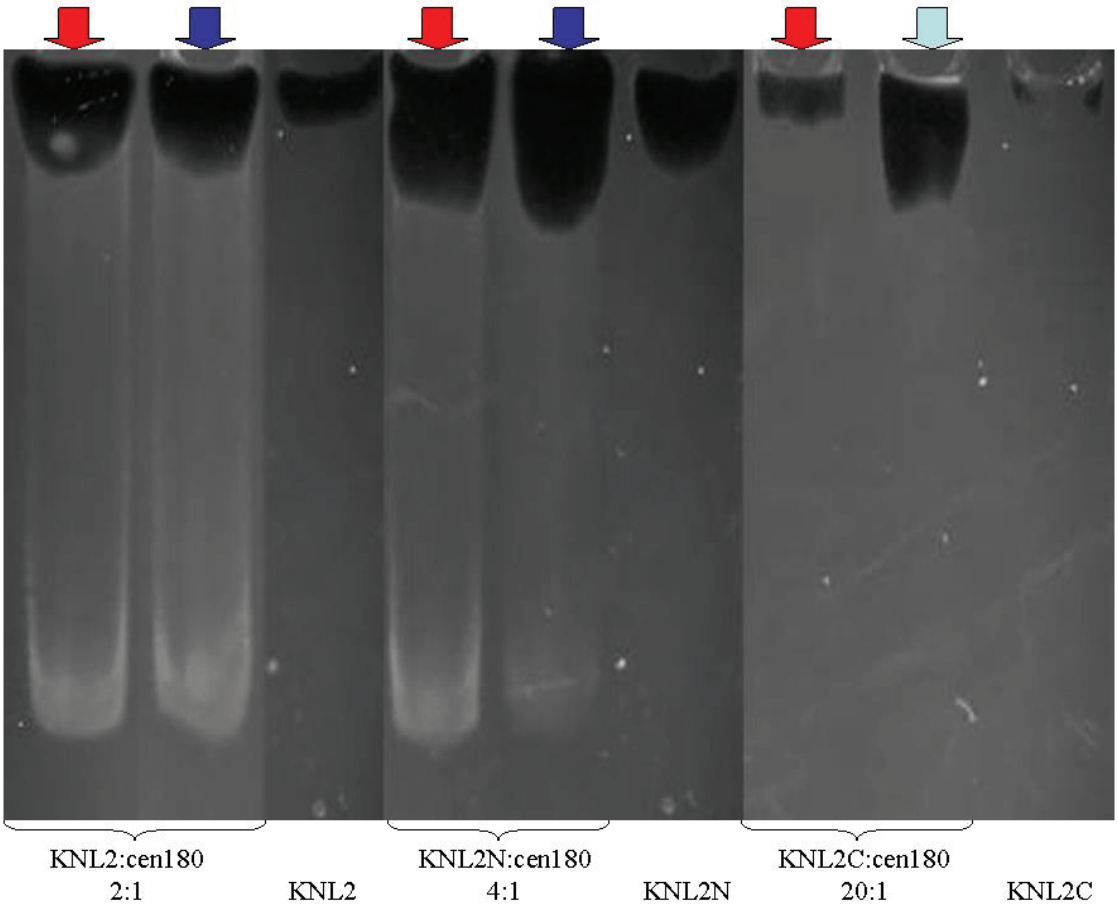
Supplemental Figure 7: Purified protein solution of the KNL2 variants were separated on a 12% SDS-PAA gel and subsequently analyzed by mass spectrometry. The designated band of KNL2 (1), KNL2Nterm (2), KNL2Cterm (3) and bands with lower molecular weight of KNL2 (4) and KNL2Nterm (5) were picked twice (a/b).

```
AT5G02520/F4KCE9      MTEPNLDEDGSKSSFQKTVVLRDWWLIKCPKEFEGKQFGVAGFEESVETRAMRVFTSSPI
gi|7413641            MTEPNLDEDGSKSSFQKTVVLRDWWLIKCPKEFEGKQFGVAGFEESVETRAMRVFTSSPI
                      ************************************************************

AT5G02520/F4KCE9      TKALDVFTLLASDGIYITLRGFLNKERVLKNGFNPEISREFIFGFPPCWERVCNSCFEGD
gi|7413641            TKALDVFTLLASDGIYITLRGFLNKERVLKNGFNPEISREFIFGFPPCWERVCNSCFEGD
                      ************************************************************

AT5G02520/F4KCE9      SFGTDVNTVPSTIEKACPPILSPCKYSNRNLKDNPAESREKSNVTETDIAEINDKGGSGA
gi|7413641            SFGTDVNTVPSTIEKACPPILSPCKYSNRNLKDNPAESREKSNVTETDIAEINDKGGSGA
                      ************************************************************

AT5G02520/F4KCE9      RDIKTARRRSLHLQIKRILESSKVRKTANDGDHGSEFLNTAKRGDVERDGCEVINNEDSE
gi|7413641            RDIKTARRRSLHLQIKRILESSKVRKTANDGDHGSEFLNTAKRGDVERDGCEVINNEDSE
                      ************************************************************

AT5G02520/F4KCE9      WKLDESEVQNLCNDGDNGSEGFIKAKSSDVEKDKSEAIDNDVISPAVGSGIKHTGADNVD
gi|7413641            WKLDESEVQNLCNDGDNGSEGFIKAKSSDVEKDKSEAIDNDVISPAVGSGIKHTGADNVD
                      ************************************************************

AT5G02520/F4KCE9      KVTSASATGESLTSEQQNGLLVTTASPHSLLKDLAKSSKPEKKGISKKSGKILRSDDNVV
gi|7413641            KVTSASATGESLTSEQQNGLLVTTASPHSLLKDLAKSSKPEKKGISKKSGKILRSDDNVV
                      ************************************************************

AT5G02520/F4KCE9      DPMNYSGTKVKSAENKRKIDASKLQSPTSNVAEHSKEGLNNAKSNDVEKDVCVAINNEVI
gi|7413641            DPMNYSGTKVKSAENKRKIDASKLQSPTSNVAEHSKEGLNNAKSNDVEKDVCVAINNEVI
                      ************************************************************

AT5G02520/F4KCE9      SPVKGFGKRLSGTDVERLTSKNATKESLTSVQRKGRVKVSKAFQDPLSKGKSKKSEKTLQ
gi|7413641            SPVKGFGKRLSGTDVERLTSKNATKESLTSVQRKGRVKVSKAFQDPLSKGKSKKSEKTLQ
                      ************************************************************

AT5G02520/F4KCE9      SNSNVVEPMNHFRSEAEEAEENLSWEKIKRKIDFDVEVTPEKKVKQQKTNAASTDSLGQK
gi|7413641            SNSNVVEPMNHFRSEAEEAEENLSWEKIKRKIDFDVEVTPEKKVKQQKTNAASTDSLGQK
                      ************************************************************

AT5G02520/F4KCE9      RSRSGRVLVSSLEFWRNQIPVYDMDRNLIQVKDGSETNSAPSKGKGSDSRKRRNLKIK
gi|7413641            RSRSGRVLVSSLEFWRNQIPVYDMDRNLIQVKDGSETNSAPSKG--------------
                      *********************************************
```

Supplemental Figure 8: The older KNL2 database identifier gi|7413641 (genbank, 2006) and the latest identifiers AT5G02520 (TAIR, 2013) or F4KCE9 (Uniprot, 2011) (Supplemental Figure 1) were analyzed by a sequence alignment tool (see 2.10). There is a difference just in the last 14 C-terminal amino acids, which are missing in the sequence of the genbank identifier.



Supplemental Figure 9: The signals of agarose gel of the EMSA (Figure 7) and of the relating chemoluminescent film (Figure 8) were merged. A small shift in the cen180 signal as well as a shift in the signal of KNL2Cterm (KNL2C) could be observed, whereas there was no overlay of DNA and protein signals.

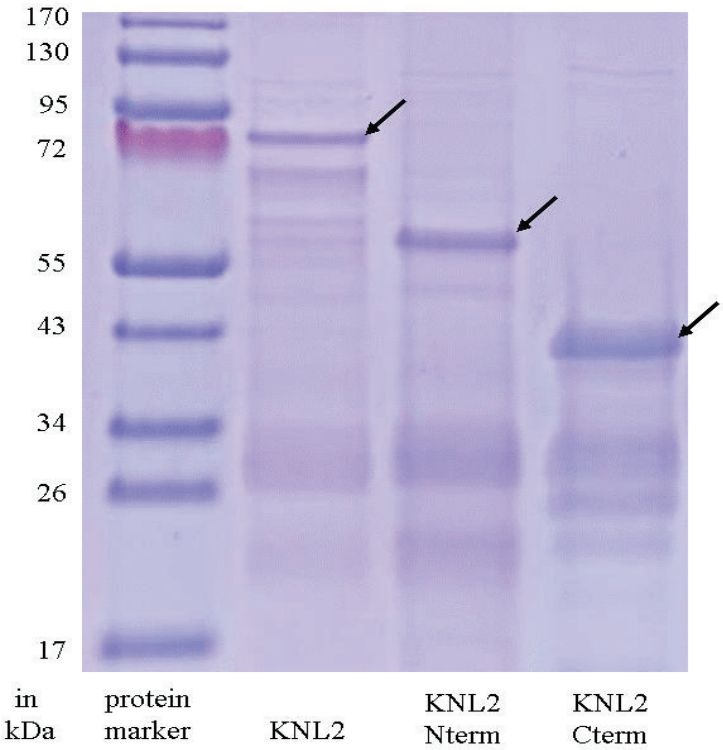


Supplemental Figure 10: A PAA gel based EMSA (see 2.22.2) was performed with the centromeric sequence cen180 and the KNL2 variants dialyzed in different concentrations of urea. DNA molecules were detected by SYBR® Safe staining and the KNL2 variants with a Western blot approach using the anti-His tag-HRP antibodies (Miltenyi Biotec®, #130-092-785, HRP coupled, 1:5000) and a chemoluminescent film, which was exposed by a triggered HRP reaction for 10 s. No clear difference could be detected between KNL2 or KNL2Nterm (KNL2N) mixed with cen180 dissolved in a concentration of 0.5 M urea (red arrowed lanes) and KNL2 or KNL2Nterm mixed with cen180 dissolved in 0.19 M urea (dark blue arrowed lanes). The band at 180 bp of cen180 disappeared and shifted near to the slots in the lanes of KNL2Cterm (KNL2C) mixed with cen180 dialyzed in 0.5 M urea (red arrowed lane) and KNL2Cterm mixed with cen180 containing no urea (light blue arrowed lane). However, the shape of the detected KNL2Cterm signal differed between both urea concentrations. Different molecular ratios were used for the different KNL2 variants according to their concentration after the purification (Figure 4).

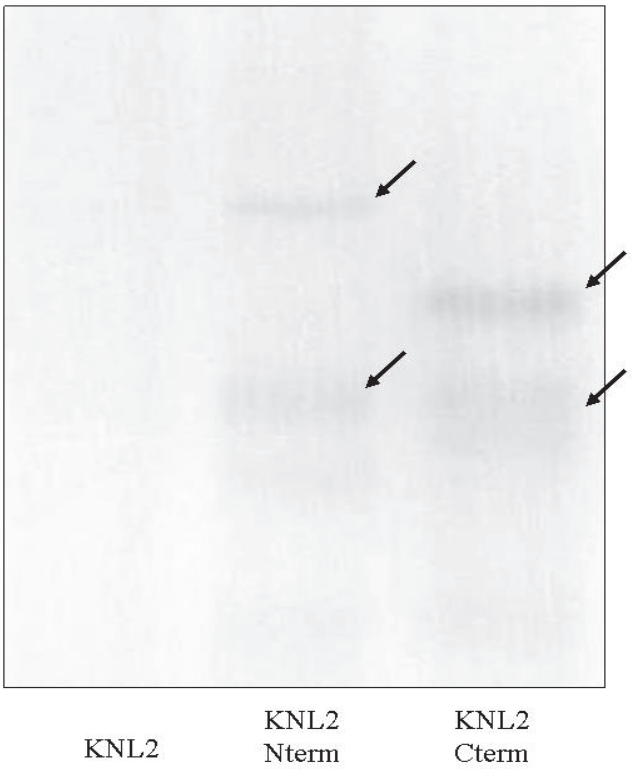
Supplemental Figure 11: The cDNA of C*enH3*, *KNL2* or *KNL2*Cterm (Supplemental Figure 1) was inserted into the plasmids pSPYCE-35s and pSPYNE-35s. These constructs were used to transform *Agrobacterium tumefaciens*. *Agrobacterium* cultures were injected into *Nicotiana benthamiana* leaves. An expression of the three proteins coupled with N-terminal (YFP-Nterm by pSPYCE-35s) or C-terminal part (YFP-Cterm by pSPYCE-35s) of YFP was induced *in planta*. The lower epidermis of injected leaves was isolated and a DIC microscopy was performed (cross-reference). The results of the microscopy are presented in Supplemental Table 1.

Supplemental Table 1: The results of the BiFC experiments (see 2.23 and Supplemental Figure 11) showed green nuclear fluorescence in any ***CenH3*** containing combination of fusion constructs, whereas *KNL2* and *KNL2*Cterm fusion constructs not.

Coupled with YFP-Nterm	Coupled with YFP-Cterm	Nuclear fluorescence Y/N?
CenH3	***CenH3***	**Y**
CenH3	Empty vector	**Y**
Empty vector	***CenH3***	**Y**
KNL2	***CenH3***	**Y**
CenH3	*KNL2*	**Y**
*KNL2*Cterm	***CenH3***	**Y**
CenH3	*KNL2*Cterm	**Y**
KNL2	*KNL2*	N
*KNL2*Cterm	*KNL2*Cterm	N
KNL2	Empty vector	N
Empty vector	*KNL2*	N
*KNL2*Cterm	Empty vector	N
Empty vector	*KNL2*Cterm	N



Supplemental Figure 12: This SDS-PAGE of the three KNL2 variants is identical to Figure 15. However, Aurora kinase 3 was not added. The KNL2 variants were loaded unequally.



Supplemental Figure 13: The SDS-PAA gel of Supplemental Figure 12 was analyzed by a phosphor-imager (see 2.25). Black arrowed bands indicate phosphorylation. The designated bands of the KNL2 variants were highlighted by arrows.

Additional N-terminal sequences (encoded by expression vector):

M S Y Y H H H H H H L E S T S L Y K K A G F

KNL2Nterm part (cDNA encoded):

M T E P N L D E D G S K S S F Q K T V V L R D W W L I K C P K E F E G K Q F
G V A G F E E S V E T R A M R V F T S S P I T K A L D V F T L L A S D G I Y
I T L R G F L N K E R V L K N G F N P E I S R E F I F G F P P C W E R V C N
S C F E G D S F G T D V N T V P S T I E K A C P P I L S P C K Y S N R N L K
D N P A E S R E K S N V T E T D I A E I N D K G G S G A R D I K T A R R R S
L H L Q I K R I L E S S K V R K T A N D G D H G S E F L N T A K R G D V E R
D G C E V I N N E D S E W K L D E S E V Q N L C N D G D N G S E G F I K A K
S S D V E K D K S E A I D N D V I S P A V G S G I K H T G A D N V D K V T S
A S A T G E S L T S E Q Q N G L L V T T A S P H S L L K D L A K S S K P E K
K G I S K K S G K I L R S D D N V V D P

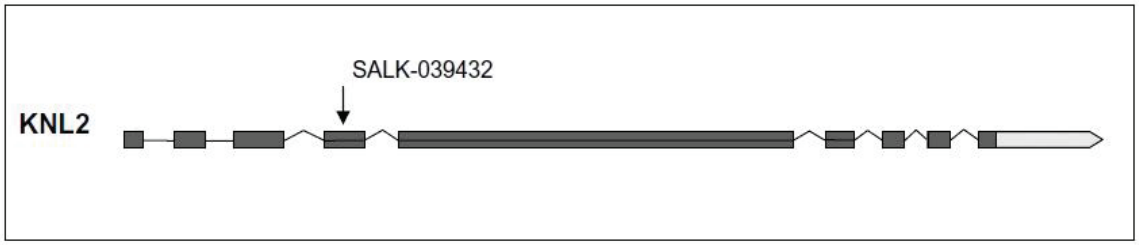
KNL2Cterm part (cDNA encoded):

M N Y S G T K V K S A E N K R K I D A S K L Q S P T S N V A E H S K E G L N
N A K S N D V E K D V C V A I N N E V I S P V K G F G K R L S G T D V E R L
T S K N A T K E S L T S V Q R K G R V K V S K A F Q D P L S K G K S K K S E
K T L Q S N S N V V E P M N H F R S E A E E A E E N L S W E K I K R K I D F
D V E V T P E K K V K Q Q K T N A A S T D S L G Q K R S R S G R V L V S S L
E F W R N Q I P V Y D M D R N L I Q V K D G S E T N S A P S K G K G S D S R
K R R N L

Additional C-terminal sequences (encoded by expression vector):

K I K D P A F L Y K V V D S R L L T K P E R K L S W L L P P L S N N **Stop**

Supplemental Figure 14: The amino acid sequence of the KNL2 variants (Supplemental Figure 1 & Table 2) were checked for conserved phosphorylation motifs of Aurora kinases. This putative conserved motif (underlined in the sequence) is R/K-R/K-X-pS/pT-(not P) (Nousiainen et al., 2006; Dephoure et al., 2008; Sardon et al., 2010). Every recombinant, expressed KNL2 variant contained the additional N-terminal and C-terminal amino acids originated by the expression vector (Supplemental Figure 1). KNL2 contained the KNL2term part and the KNL2Cterm part. There are 43 aspartic acid residues (green) and 12 histidine residues (violet) spread over the whole sequences. These might be canonical target residues for *E. coli* kinases. Additionally 76 serine (red) and 34 threonine (blue) residues were found, which can be phosphorylated too (Pereira et al., 2011). The last five C-terminal amino acids contain two non-polar amino acids: leucine and proline.



Supplemental Figure 15: The exon- (grey bars) intron- (lines between exons) structure of the *KNL2* gene is shown. The term '*KNL2* SALK mutant' is referring to a TDNA insertion at exon 4 with the shown accession number. This figure is part of the publication of Lermontova et al. (2013)

Supplemental Table 2: The sequences of the recombinant KNL2 variant (Supplemental Figure 14) were analyzed *in silico* (http://web.expasy.org/compute_pi/) to obtain the values for the isoelectric point (pI). CenH3 has a pI value of 11.58.

Protein	pI
KNL2	8.96
KNL2Nterm	6.00
KNL2Cterm	9.86

Supplemental Table 3: Certain codons, which are less used by *E. coli* (Kane, 1995; Green and Sambrook, 2012), were counted in the cDNA sequence of the *KNL2* variants (see Supplemental Figure 1) by using *in silico* tools (http://www.kazusa.or.jp/codon/countcodon.html).

	KNL2	*KNL2*Nterm	*KNL2*Cterm
AGG	11	6	7
AGA	12	8	4
AUA	8	6	2
CUA	2	1	2
CCC	4	3	2
GGA	17	11	6
Sum of raw codons	54	35	23
Total codon number	652	416	290
ratio	8,3%	8,4%	7,9%

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und benutzte Werke kenntlich gemacht habe.